Simple, Robust, and Accurate F and t Tests in Cointegrated Systems

Jungbin Hwang
Department of Economics,
University of Connecticut

Yixiao Sun*
Department of Economics,
University of California, San Diego

First version: July 2014. This version: July 2017

Abstract

This paper proposes new, simple, and more accurate statistical tests in a cointegrated system that allows for endogenous regressors and serially dependent errors. The approach involves first transforming the time series using orthonormal basis functions in $L^2[0,1]$, which has energy concentrated at low frequencies, and then running an augmented regression based on the transformed data and constructing the test statistics in the usual way. The approach is essentially the same as the trend instrumental variable approach of Phillips (2014), but we hold the number of orthonormal basis functions fixed in order to develop the standard F and t asymptotic theory. The tests are extremely simple to implement, as they can be carried out in exactly the same way as if the transformed regression is a classical linear normal regression. In particular, critical values are from the standard F or t distribution. The proposed F and t tests are robust in that they are asymptotically valid regardless of whether the number of basis functions is held fixed or allowed to grow with the sample size. The F and t tests have more accurate size in finite samples than existing tests such as the asymptotic chi-squared and normal tests based on the fully modified OLS estimator of Phillips and Hansen (1990) and can be made as powerful as the latter test.

JEL Classification: C12, C13, C32

Keywords: Cointegration, F test, Alternative Asymptotics, Nonparametric Series Method, t test, Transformed and Augmented OLS

1 Introduction

This paper considers a new approach to statistical inference in a triangular cointegrated regression system. A salient feature of this system is that the I(1) regressors are endogenous. In addition, to maintain generality of the short run dynamics, we allow the I(0) regression errors to have serial dependence of unknown forms. One of the most popular semiparametric estimators

*Email: jungbin.hwang@uconn.edu and yisun@ucsd.edu. Correspondence to: Yixiao Sun, Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. We thank Michael Jansson and two anonymous referees for constructive comments. We are particularly grateful to Peter C. B. Phillips, the editor, for his detailed comments, constructive suggestions, and considerable input. We acknowledge partial research support from NSF under Grant No. SES-1530592.

in this system is the fully modified OLS (FMOLS) estimator of Phillips and Hansen (1990). The estimator involves using a long run variance and a half long run variance to remove the long run joint dependence and endogeneity bias. Both the long run variance and the half long run variance are estimated nonparametrically. Inference based on the FMOLS is standard — the Wald statistic is asymptotically chi-squared as in the classical linear regression with stationary or iid data. This is perhaps one of the most elegant and convenient results in time series econometrics. It releases us from having to simulate functionals of Brownian motion.

One drawback of the FMOLS method is that the asymptotic chi-squared test often has large size distortion. The source of the problem is that the estimation errors in the long run variance and half long run variance have been completely ignored in the conventional asymptotic framework adopted in Phillips and Hansen (1990). A new "fixed-b" asymptotic framework has been put forward by Vogelsang and Wagner (2014), but the Wald statistic does not appear to be asymptotically pivotal, making inference difficult and inconvenient. For this reason, Vogelsang and Wagner (2014) proceed to propose a different estimation method called the Integrated Modified OLS (IMOLS). They show that the associated test statistics are asymptotically pivotal under fixed-b asymptotics. However, the limiting distributions are nonstandard, and critical values have to be simulated.

In the same spirit of Vogelsang and Wagner (2014), we consider an alternative estimation method that involves first transforming the data using orthonormal basis functions and then running an augmented regression based on the transformed data in the second stage. This gives rise to our transformed and augmented (TA) OLS (TAOLS) estimator. Augmentation removes the long run dependence problem, and transformation eliminates the second-order bias that plagues the OLS estimator.

Our TAOLS estimator is closely related to the trend instrument variable (TIV) estimator of Phillips (2014). Phillips (2014) considers the augmented regression model in the time domain and runs an instrumental variable regression using orthonormal basis functions as instruments. Depending on the trend functions used, the TAOLS and TIV estimators may be numerically identical or asymptotically equivalent under the asymptotics considered in this paper. In essence, the two estimators extract the signals on long run comovements in an identical way — both involve projecting the underlying time series on a set of orthonormal deterministic basis functions.

A key feature of our asymptotic analysis is that the number of basis functions K is held fixed as the sample size goes to infinity, leading to our fixed-K asymptotic theory. Compared with existing methods such as the FMOLS of Phillips and Hansen (1990) and the IMOLS of Vogelsang and Wagner (2014), our new method enjoys several advantages.

First, under the fixed-K asymptotics, the test statistics based on the TAOLS estimator are asymptotically standard F or t distributed. Since the critical values from the F and t distributions are easily available from statistical tables, there is no need to further approximate or simulate nonstandard limit distributions. In addition, the test statistics can be obtained directly from statistical programs that can compute the F and t statistics in a classical linear normal regression. So, our method is practically convenient and empirically appealing in comparison to the IMOLS method, where the fixed-b limiting distribution is highly nonstandard and the critical values have to be simulated.

Second, given that the TAOLS estimator is asymptotically equivalent to the TIV estimator of Phillips (2014), we have also established the fixed-K asymptotics of the TIV estimator and the associated test statistics. Under the increasing-K asymptotics where K grows with the sample size at an appropriate rate, Phillips (2014) shows that the Wald statistic and the t statistic are asymptotically chi-squared and normal, respectively. While the fixed-K asymptotic distribution is different from the increasing-K asymptotic distribution, we show that the fixed-K asymptotic distribution approaches the increasing-K asymptotic distribution as K increases. As a result, the fixed-K critical values are asymptotically valid regardless of the type of asymptotics we consider. This is a robust property enjoyed by our asymptotic F and t tests.

Third, our simulation results show that the asymptotic F and t tests have more accurate size than existing tests such as the asymptotic chi-squared and normal tests based on the FMOLS estimator. By choosing K appropriately, the asymptotic F and t tests can be as powerful as the latter tests. This is based on our simulation evidence. It is also consistent with the asymptotic efficiency of the TAOLS estimator under the increasing-K asymptotics. The asymptotic efficiency holds because the TAOLS estimator and the asymptotically efficient FMOLS estimator have the same asymptotic distribution under the increasing-K asymptotics.

Finally, taking it literally, the fixed-K asymptotics require us to use only low-frequency information. Fundamentally, what a cointegrating vector measures is the long run relation among economic time series. For this reason, it is natural to estimate the cointegrating vector using only the long run variation of the underlying time series. Doing so helps us avoid high-frequency contaminations. From this perspective, the fixed-K limiting thought experiment not only is an asymptotic device for developing new and more accurate approximations but also has substantive empirical content in economic applications.

This paper contributes to a large body of literature on semiparametric estimation of cointegrated systems with Phillips and Hansen (1990), Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993) as seminal early contributions. In the FMOLS setting, partial fixed-b asymptotic theory for cointegration inference has been considered by Bunzel (2006) and Jin, Phillips, and Sun (2006), but the fixed-b asymptotics is applied only to the standard error estimator; see Vogelsang and Wagner (2014) for further discussion of this. Relative to the TIV approach of Phillips (2014), the contributions of the current paper lie more in the fixed-K asymptotics than in the proposed TAOLS estimator, as the TIV estimator and the TAOLS estimator turn out to be essentially the same.

Transforming a time series using the basis functions considered in this paper is equivalent to filtering the time series with a particular class of linear filters. The filtering idea has a long history; see, for example, the seminal contribution of Thomson (1982). For a textbook treatment, see Chapter 5 of Stoica and Moses (2005). This idea has been used in nonparametric cointegration analysis. Bierens (1997) and Müller and Watson (2013) employ basis-function transforms in order to extract the long run variation and covariation in the underlying time series. Without imposing a parametric VAR structure as in Johansen (1991), Bierens (1997) proposes nonparametric tests for the number of cointegrating vectors, which is the same as the degree of rank deficiency of a standardized long run variance matrix. Bierens's test statistics involve functions of the eigenvalues of this long run variance matrix and have nonstandard limiting distributions. A variant of Bierens's method appears in Shintani (2001) who employs kernel long run variance estimators instead of series long run variance estimators. This idea of using rank deficiency to test for the cointegration rank can be traced to Phillips and Ouliaris (1990). Müller and Watson (2013) use the Neyman-Pearson decision-theoretic framework to design robust and nearly optimal tests about the cointegrating vectors when they are fully specified under the null hypothesis, that is, when all the cointegrating vectors are known under the null hypothesis. Both Bierens (1997) and Müller and Watson (2013) consider a fixed number of basis functions, which is in the same spirit as the fixed-K asymptotics we consider here. However, our paper has different objectives:

our aim is to estimate the cointegrating vector and to conduct inferences on both the full vector and a subvector.

Basis-function transformations have also been used in heteroskedasticity and autocorrelation robust (HAR) inference. The most recent research along this line was inspired by Phillips (2005b), although the idea can be traced back to the much earlier literature on the multiple window method for spectral estimation, started by Thomson (1982). The term "HAR" was first introduced by Phillips (2005a). While the current paper employs basis-function transformation as a tool to estimate the main parameters of interest, the HAR literature uses them to estimate the asymptotic variances of parameter estimators, leading to the class of orthonormal series HAR variance estimators. Using this type of variance estimators, F and t limit theory has been established in the HAR literature. See Sun (2011) for trend regressions, Sun (2013) for stationary moment processes, and Sun (2014c) for highly persistent moment processes. Sun and Kim (2012) develop the F or t approximation to the J statistic, while Sun and Kim (2015) develop F and t limit theory in a spatial setting. Hwang and Sun (2017) develop the F or t limit theory in a two-step generalized method of moments (GMM) framework with a series HAR variance estimator used as the weighting matrix. This paper complements the F and t limit theory established in these papers. Readers are referred to Müller and Watson (2016) for further discussion on basis-function transformations and their applications in econometrics.

The rest of the paper is organized as follows. Section 2 introduces a standard linear cointegration regression and discusses some of the drawbacks of existing methods. Section 3 introduces our TAOLS estimator and establishes the fixed-K asymptotic limits of the TAOLS estimator and the corresponding Wald statistic. Section 4 considers cointegration analysis under half cosine or shifted cosine transforms. Section 5 presents simulation evidence. The last section concludes. Proofs are given in the appendix.

2 Model and Existing Literature

Following Vogelsang and Wagner (2014), we consider the cointegration model

$$\begin{aligned} y_t &= \alpha_0 + x_t'\beta_0 + u_{0t} \\ x_t &= x_{t-1} + u_{xt} \end{aligned} \tag{1}$$

for $t = 1, \ldots, T$, where y_t is a scalar time series and x_t is a $d \times 1$ vector of time series with $x_0 = O_p(1)$. The mean-zero error vector $u_t \equiv (u_{0t}, u_{xt}')' \in \mathbb{R}^m$ for $m = d + 1$ is jointly stationary with long run variance (LRV) matrix Ω. We partition Ω as

$$\underset{m\times m}{\Omega} = \sum_{j=-\infty}^{\infty} Eu_t u_{t-j}' = \begin{pmatrix} \underset{1\times 1}{\sigma_0^2} & \underset{1\times d}{\sigma_{0x}} \\ \underset{d\times 1}{\sigma_{x0}} & \underset{d\times d}{\Omega_{xx}} \end{pmatrix}, \tag{2}$$

and write it as a sum of three conformable components: $\Omega = \Sigma + \Lambda + \Lambda'$, where

$$\Lambda := \sum_{j=1}^{\infty} Eu_{t-j}u_t' = \begin{pmatrix} \underset{1\times 1}{\Lambda_{00}} & \underset{1\times d}{\Lambda_{0x}} \\ \underset{d\times 1}{\Lambda_{x0}} & \underset{d\times d}{\Lambda_{xx}} \end{pmatrix} \text{ and } \Sigma := Eu_t u_t' = \begin{pmatrix} \underset{1\times 1}{\Sigma_{00}} & \underset{1\times d}{\Sigma_{0x}} \\ \underset{d\times 1}{\Sigma_{x0}} & \underset{d\times d}{\Sigma_{xx}} \end{pmatrix}.$$

The half long run variance Δ is defined to be

$$\Delta = \Sigma + \Lambda = \begin{pmatrix} \Delta_{00} & \Delta_{0x} \\ \Delta_{x0} & \Delta_{xx} \end{pmatrix}. \tag{3}$$

We assume that Ω_{xx} is positive definite so that x_t is a full-rank integrated process.

We shall assume the Functional Central Limit Theorem (FCLT):

$$T^{-1/2}\sum_{t=1}^{[T\cdot]} u_t \Rightarrow B(\cdot) = \Omega^{1/2}W(\cdot), \tag{4}$$

where $W(\cdot) := (w_0(\cdot), W_x'(\cdot))'$ is an m-dimensional standard Brownian process. Also, it will be convenient in our asymptotic development to represent the process $B(\cdot)$ using the Cholesky form of $\Omega^{1/2}$:

$$B(\cdot) = \begin{pmatrix} B_0(\cdot) \\ B_x(\cdot) \end{pmatrix} = \begin{pmatrix} \sigma_{0\cdot x} w_0(\cdot) + \sigma_{0x}\Omega_{xx}^{-1/2} W_x(\cdot) \\ \Omega_{xx}^{1/2} W_x(\cdot) \end{pmatrix}, \tag{5}$$

where $\sigma_{0\cdot x}^2 = \sigma_0^2 - \sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}$ and $\Omega_{xx}^{1/2}$ is a symmetric matrix square root of Ω_{xx}.

To simplify the discussion, we assume that there is no intercept in the regression. Let $X = [x_1', \ldots, x_T']'$ and $Y = [y_1, \ldots, y_T]'$. The OLS estimator of β_0 is given by $\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$. It follows from Phillips and Durlauf (1986) and Stock (1987) that

$$T(\hat{\beta}_{OLS} - \beta_0) = \left(\frac{1}{T^2}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} x_t u_{0t}\right) \tag{6}$$

$$\Rightarrow \left(\int_0^1 B_x(r) B_x'(r) dr\right)^{-1} \left(\int_0^1 B_x(r) dB_0(r) + \Delta_{x0}\right), \tag{7}$$

where Δ_{x0} reflects the second-order endogeneity bias.

Since $B_x(\cdot)$ and $B_0(\cdot)$ are correlated, and Δ and hence Δ_{x0} are unknown, it is not possible to make an asymptotically valid inference based on the naive OLS estimator. To overcome these two problems (correlation $B_x(\cdot)$ and $B_0(\cdot)$ and second-order endogeneity bias), Phillips and Hansen (1990) suggest the FMOLS method that involves estimating Ω and Δ in the first step. Typical estimators of Ω and Δ take the following forms:

$$\hat{\Omega} = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} Q_h\left(\frac{s}{T}, \frac{t}{T}\right)\hat{u}_t \hat{u}_s', \tag{8}$$

$$\hat{\Delta} = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=s}^{T} Q_h\left(\frac{s}{T}, \frac{t}{T}\right)\hat{u}_t \hat{u}_s', \tag{9}$$

where $\hat{u}_t = (\hat{u}_{0t}, u_{xt}')'$ and $\hat{u}_{0t} = y_t - x_t'\hat{\beta}_{OLS}$. In the above definitions of $\hat{\Omega}$ and $\hat{\Delta}$, $Q_h(r,s)$ is a symmetric weighting function that depends on the smoothing parameter h. For conventional kernel LRV estimators, $Q_h(r,s) = k((r-s)/b)$ and we take $h = 1/b$. For orthonormal series (OS) LRV estimators, $Q_h(r,s) := Q_K(r,s) = K^{-1}\sum_{j=1}^{K}\phi_j(r)\phi_j(s)$ and we take $h = K$, where $\left\{\phi_j(r)\right\}_{j=1}^{K}$ are orthonormal basis functions in $L^2[0,1]$ satisfying $\int_0^1 \phi_j(r)\, dr = 0$. For more background information on OS LRV estimation, see Sun (2011, 2013). We parametrize h in such a way that h indicates the amount of smoothing for both types of LRV estimators.

After partitioning $\hat{\Omega}$ and $\hat{\Delta}$ in the same way as Ω and Δ, we define

$$\begin{aligned} y_t^+ &:= y_t - \hat{\sigma}_{0x}\hat{\Omega}_{xx}^{-1}\Delta x_t, \\ u_t^+ &:= u_t - \hat{\sigma}_{0x}\hat{\Omega}_{xx}^{-1}\Delta x_t, \\ \mathcal{M} &:= T(\hat{\Delta}_{x0} - \hat{\Delta}_{xx}\hat{\Omega}_{xx}^{-1}\hat{\sigma}_{x0}). \end{aligned} \tag{10}$$

Then the FMOLS estimator is given by

$$\hat{\beta}_{FM} = \left(X'X\right)^{-1}\left(X'Y^{+} - \mathcal{M}\right), \tag{11}$$

where $Y^{+} = [y_1^{+}, \ldots, y_T^{+}]'$. On the basis of the kernel estimators of Ω and Δ, Phillips and Hansen (1990) show that $\hat{\beta}_{FM}$ is asymptotically mixed normal, that is,

$$T(\hat{\beta}_{FM} - \beta_0) \Rightarrow MN\left(0, \sigma_{0\cdot x}^2\left[\int_0^1 B_x(r)B_x'(r)dr\right]^{-1}\right). \tag{12}$$

This is in contrast to the limiting distribution of $\hat{\beta}_{OLS}$, which is complicated and has a second-order endogeneity bias. Based on a consistent estimator $\hat{\sigma}_{0\cdot x}^2$ of $\sigma_{0\cdot x}^2$, one can obtain t and Wald statistics that are asymptotically normal and chi-squared distributed, respectively.

A key step behind Phillips and Hansen's result is that $\hat{\Omega}$, $\hat{\Delta}$, and $\hat{\sigma}_{0\cdot x}^2$ are all approximated by the respective degenerate distributions concentrated at Ω, Δ, and $\sigma_{0\cdot x}^2$. That is, regardless of the kernel function and the bandwidth used in the nonparametric estimators $\hat{\Omega}$, $\hat{\Delta}$, and $\hat{\sigma}_{0\cdot x}^2$, the same asymptotic approximations are used. However, in finite samples, both the kernel function and the bandwidth, especially the latter, do affect the sampling distribution of $\hat{\beta}_{FM}$ and the associated test statistics. For this reason, the normal and chi-squared approximations can be very poor in finite samples. This is because these approximations completely ignore the estimation uncertainty in the nonparametric estimators $\hat{\Omega}$, $\hat{\Delta}$, and $\hat{\sigma}_{0\cdot x}^2$, which can be very high in finite samples. Bunzel (2006) and Jin, Phillips, and Sun (2006) develop partial fixed-b asymptotic theory that accounts for the estimation uncertainty in $\hat{\sigma}_{0\cdot x}^2$ but ignores that in $\hat{\Omega}$ and $\hat{\Delta}$.

The degenerate distributional approximations for $\hat{\Omega}$, $\hat{\Delta}$, and $\hat{\sigma}_{0\cdot x}^2$ with consequential normal and chi-squared tests are obtained under the conventional increasing-smoothing asymptotic theory. Instead of the conventional asymptotics, we can use the alternative fixed-smoothing asymptotics to obtain more accurate asymptotic approximations. The fixed-smoothing asymptotics include the fixed-b asymptotics of Kiefer and Vogelsang (2005) as a special case. For further discussion of these two types of asymptotics, see Sun (2014a, 2014b). There is a growing number of papers on fixed-b asymptotic theory for stationary data starting with Kiefer and Vogelsang (2005). More recently, Vogelsang and Wagner (2014) develop a fully-fledged fixed-b asymptotic theory for the FMOLS estimator and show that when the estimation uncertainty in $\hat{\Omega}$ and $\hat{\Delta}$ is accounted for, the FMOLS estimator still suffers from a second-order asymptotic bias and has an asymptotic variance that is much more complex than that given by Phillips and Hansen (1990). As a result, the Wald and t statistics depend on many nuisance parameters even in the limit, and this makes the fixed-b asymptotic theory hard to use.

As an alternative solution, Vogelsang and Wagner (2014) suggest the Integrated Modified OLS (IMOLS) estimator, which is based on partial sums of the original cointegrating regression augmented by the original regressor. They invoke the fixed-b asymptotics to approximate the IMOLS test statistics and show that they are asymptotically pivotal. However, their limiting distributions are highly nonstandard; the critical values have to be simulated for practical implementation.

3 Cointegration Analysis: Augmentation and Transformation

3.1 Model without time trend

To confront several challenges in the literature, we propose an alternative method to estimate the cointegration model in (1) where no trend is present. We follow Phillips (2014) and consider the augmented cointegration model

$$y_t = \alpha_0 + x_t'\beta_0 + \Delta x_t'\delta_0 + u_{0\cdot xt}, \tag{13}$$

where $\delta_0 = \Omega_{xx}^{-1}\sigma_{x0}$ is the long run regression coefficient of Δx_t on u_{0t}, and $u_{0\cdot xt} = u_{0t} - \delta_0' u_{xt}$ is the long run regression error of u_{0t} projected onto u_{xt}. The long run variance of $u_{0\cdot xt}$ is $\sigma_{0\cdot x}^2$.

Let $\{\phi_i\}_{i=1}^{\infty}$ be a set of orthonormal basis functions in the standard Hilbert space $L^2[0,1]$. Our method starts by transforming the original data $\{y_t, x_t', \Delta x_t'\}_{t=1}^T$ using the basis functions $\{\phi_i\}_{i=1}^K$ for a finite K and then conducts a regression analysis based on the transformed data. For each $i = 1, \ldots, K$, the transformed data $\{\mathbb{W}_i\}$ are weighted averages of the original data:

$$\begin{aligned}
\mathbb{W}_i^{\alpha} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i(\frac{t}{T}),\\
\mathbb{W}_i^{y} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_t\phi_i(\frac{t}{T}) = \frac{Y'\Phi_i}{\sqrt{T}},\ \mathbb{W}_i^{x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t\phi_i(\frac{t}{T}) = \frac{X'\Phi_i}{\sqrt{T}},\\
\mathbb{W}_i^{\Delta x} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta x_t\phi_i(\frac{t}{T}),\ \mathbb{W}_i^{0\cdot x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{0\cdot xt}\phi_i(\frac{t}{T}),
\end{aligned} \tag{14}$$

where $\Phi_i = [\phi_i(1/T), \ldots, \phi_i((T-1)/T), \phi_i(1)]'$ is the basis vector corresponding to the basis function $\phi_i(\cdot)$. In the context of (realized) variance estimation, such a transform has been used in Phillips (2005b), Sun (2006), and Müller (2007), among others.

When $\phi_i(r) = \phi_i(1-r)$, which holds for the Fourier basis functions we use, we can write, for example,

$$\mathbb{W}_i^{y} = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}y_{T-t}\phi_i(\frac{T-t}{T}) = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}y_{T-t}\phi_i(\frac{t}{T}). \tag{15}$$

Therefore, $\mathbb{W}_i^y$ can be regarded as the output from applying a linear filter to $\{y_t\}_{t=1}^T$. The transfer function of this linear filter is

$$H_{Ti}(\omega) = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}\phi_i(\frac{t}{T})\exp(\iota t\omega) \text{ for } \iota = \sqrt{-1}. \tag{16}$$

To capture the long run behavior of the processes, we implicitly require that $H_{Ti}(\omega)$ be concentrated around the origin. That is, $H_{Ti}(\omega)$ resembles a band-pass filter that passes low frequencies within a certain range and attenuates frequencies outside that range. This requirement can be met by any low-order trigonometric basis such as $\sqrt{2}\sin(2\pi ir)$, $\sqrt{2}\cos(2\pi ir)$ for small i. In fact, the transfer functions associated with the first few basis functions in a commonly-used base system in $L^2[0,1]$ are often concentrated around the origin, so the requirement can be easily met.

Based on the augmented regression and the transformed data, we have

$$\mathbb{W}_i^{y} = \alpha_0\mathbb{W}_i^{\alpha} + \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x} \text{ for } i = 1, \ldots, K. \tag{17}$$

This can be regarded as a cross-sectional regression with K observations. We assume that K is large enough so that the number of observations is not smaller than the number of non-diminishing regressors. Obviously, there is no point in considering $K > T$, as there is no extra information beyond the first T transforms.

Under the assumption that each function $\phi_i(\cdot)$ is continuously differentiable and satisfies $\int_0^1 \phi_i(r)\,dr = 0$, which we will maintain, we have

$$\mathbb{W}_i^{\alpha} = \sqrt{T}\int_0^1 \phi_i(r)dr + \sqrt{T}O(1/T) = O(1/\sqrt{T}) = o(1), \tag{18}$$

and so the effect of the constant term α_0 in (17) is asymptotically negligible for a large T. As a result, our asymptotic theory remains the same regardless of whether an intercept is present. To simplify the presentation, we will assume without loss of generality that there is no intercept in the model so that

$$y_t = x_t'\beta_0 + u_{0t},\ \ x_t = x_{t-1} + u_{xt} \tag{19}$$

and

$$\mathbb{W}_i^{y} = \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x} \text{ for } i = 1,\ldots,K. \tag{20}$$

Putting (20) in vector form, we have

$$\mathbb{W}^{y} = \mathbb{W}^{x}\beta_0 + \mathbb{W}^{\Delta x}\delta_0 + \mathbb{W}^{0\cdot x}, \tag{21}$$

where $\mathbb{W}^y = (\mathbb{W}_1^y,\ldots,\mathbb{W}_K^y)'$ and $\mathbb{W}^x$, $\mathbb{W}^{\Delta x}$, and $\mathbb{W}^{0\cdot x}$ are defined similarly. Running OLS based on the above equation leads to our transformed and augmented OLS (TAOLS) estimator of $\gamma_0 = (\beta_0', \delta_0')'$:

$$\hat{\gamma}_{TAOLS} = (\tilde{\mathbb{W}}'\tilde{\mathbb{W}})^{-1}\tilde{\mathbb{W}}'\mathbb{W}^y,$$

where $\tilde{\mathbb{W}} = \left(\mathbb{W}^x, \mathbb{W}^{\Delta x}\right)$.

The TAOLS approach is closely related to the trend instrumental variable (TIV) approach of Phillips (2014), which involves solving

$$\begin{aligned}(\hat{\beta}_{IV}', \hat{\delta}_{IV}') &= \underset{(\beta,\delta)}{\arg\min}(Y - X\beta - \Delta X\delta)'\Phi(\Phi'\Phi)^{-1}\Phi'(Y - X\beta - \Delta X\delta) \\ &= (\tilde{X}'P_\Phi\tilde{X}')^{-1}(\tilde{X}'P_\Phi Y), \text{ where } \tilde{X} = [X, \Delta X] \text{ and } P_K = \Phi(\Phi'\Phi)^{-1}\Phi'. \end{aligned} \tag{22}$$

The basis functions $\Phi = [\Phi_1, ..., \Phi_K]$ now play the role of instruments for the augmented regression given in (13). When $\Phi'\Phi = I_K$, the criterion function in (22) becomes $\left\|\Phi'(Y - X\beta - \Delta X\delta)\right\|^2$, which is the same as $\left\|\mathbb{W}^y - \mathbb{W}^x\beta + \mathbb{W}^{\Delta x}\delta\right\|^2$, the sum of the squared residuals based on the transformed and augmented regression in (21). Therefore, when $\Phi'\Phi = I_K$, the TIV estimator is numerically identical to the TAOLS estimator. We can show that when K is fixed and $\int_0^1 \phi_i(r)\phi_j(r)\,ds = 1\{i = j\}$, the two estimators are asymptotically equivalent; see Proposition 2.

Let

$$P_x = \mathbb{W}^x(\mathbb{W}^{x\prime}\mathbb{W}^x)^{-1}\mathbb{W}^{x\prime},\ \ P_{\Delta x} = \mathbb{W}^{\Delta x}(\mathbb{W}^{\Delta x\prime}\mathbb{W}^{\Delta x})^{-1}\mathbb{W}^{\Delta x\prime},$$

and $M_x = I_K - P_x$, $M_{\Delta x} = I_K - P_{\Delta x}$. Then we can represent $\hat{\gamma}_{TAOLS}$ as

$$\hat{\gamma}_{TAOLS} = \begin{pmatrix} \hat{\beta}_{TAOLS} \\ \hat{\delta}_{TAOLS} \end{pmatrix} = \begin{pmatrix} (\mathbb{W}^{x\prime}M_{\Delta x}\mathbb{W}^x)^{-1}(\mathbb{W}^{x\prime}M_{\Delta x}\mathbb{W}^y) \\ (\mathbb{W}^{\Delta x\prime}M_x\mathbb{W}^{\Delta x})^{-1}(\mathbb{W}^{\Delta x\prime}M_x\mathbb{W}^y) \end{pmatrix}. \tag{23}$$

To establish the asymptotic properties of $\hat{\gamma}_{TAOLS}$, we make the following assumptions.

Assumption 1 *(i) For $i = 1,\ldots,K$, each function $\phi_i(\cdot)$ is continuously differentiable; (ii) for $i = 1,\ldots,K$, each function $\phi_i(\cdot)$ satisfies $\int_0^1 \phi_i(x)\,dx = 0$; (iii) the functions $\{\phi_i(\cdot)\}_{i=1}^K$ are orthonormal in $L^2[0,1]$.*

Assumption 2 *The vector process* $\{u_t = (u_{0t}, u'_{xt})'\}_{t=1}^T$ *satisfies the FCLT in* (4).

Assumption 1 is mild and is satisfied by many basis functions. For example, the Fourier bases $\sqrt{2}\cos(2\pi ir)$ and $\sqrt{2}\sin(2\pi ir)$ satisfy Assumption 1. Assumption 2 is a standard FCLT for time series data.

Under Assumptions 1(i) and 2, we have

$$\begin{aligned}\mathbb{W}_i^{0\cdot x} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i(\frac{t}{T})u_{0\cdot xt} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i(\frac{t}{T})\left(u_{0t}-\delta_0' u_{xt}\right)\\ &\Rightarrow \int_0^1 \phi_i(r)\,d\left[B_0(r)-\delta_0' B_x(r)\right] = \sigma_{0\cdot x}\int_0^1 \phi_i(r)\,dw_0(r)\end{aligned}$$

using the representation in (5). Here the weak convergence follows from summation by parts, the continuous mapping theorem, and integration by parts. Similarly,

$$\mathbb{W}_i^{\Delta x} \Rightarrow \int_0^1 \phi_i(r)\,dB_x(r) = \Omega_{xx}^{1/2}\int_0^1 \phi_i(r)\,dW_x(r).$$

Invoking the continuous mapping theorem again, we have

$$\frac{\mathbb{W}_i^x}{T} = \frac{1}{T^{3/2}}\sum_{s=1}^{T}\phi_i\left(\frac{s}{T}\right)x_s = \frac{1}{T}\sum_{s=1}^{T}\phi_i\left(\frac{s}{T}\right)\frac{1}{\sqrt{T}}\sum_{\tau=1}^{s}u_{x\tau} + o_p(1) \Rightarrow \Omega_{xx}^{1/2}\int_0^1 \phi_i(r)\,W_x(r)dr,$$

where the $o_p(1)$ term follows from the assumption that $x_0 = O_p(1)$ and Assumption 1(ii).

To provide some intuition, we let

$$\nu_i = \int_0^1 \phi_i(r)\,dw_0(r),\ \xi_i = \int_0^1 \phi_i(r)\,dW_x(r),\ \text{and } \eta_i = \int_0^1 \phi_i(r)\,W_x(r)dr.$$

Then the TA regression in (20) can be regarded as asymptotically equivalent to the pseudo-regression

$$\mathbb{W}_i^y \approx \eta_i'\Omega_{xx}^{1/2}(T\beta_0) + \xi_i'\Omega_{xx}^{1/2}\delta_0 + \sigma_{0\cdot x}\nu_i \text{ for } i = 1,\ldots,K. \tag{24}$$

Because ν_i is a functional of $w_0(\cdot)$, ξ_j and η_j are functionals of $W_x(\cdot)$, and $w_0(\cdot)$ is independent of $W_x(\cdot)$, the error term $\sigma_{0\cdot x}\nu_i$ is independent of the regressors $\{\eta_j'\Omega_{xx}^{1/2}, j = 1,\ldots,K\}$ and $\{\xi_j'\Omega_{xx}^{1/2}, j = 1,\ldots,K\}$. More importantly, Assumption 1(iii) ensures that $\sigma_{0\cdot x}\nu_i$ is iid normal $N(0,\sigma_{0\cdot x}^2)$. So the TA regression resembles a classical linear normal regression.

Let

$$\begin{aligned}\nu &\equiv (\nu_1,\nu_2,\ldots,\nu_K)' \in \mathbb{R}^{K\times 1},\\ \xi &\equiv (\xi_1,\xi_2,\ldots,\xi_K)' \in \mathbb{R}^{K\times d},\\ \eta &\equiv (\eta_1,\eta_2,\ldots,\eta_K)' \in \mathbb{R}^{K\times d},\end{aligned}$$

$$\tilde{\zeta} = \left(\eta\Omega_{xx}^{1/2}, \xi\Omega_{xx}^{1/2}\right),\ \tilde{\nu} = \sigma_{0\cdot x}\nu,$$

and

$$\Upsilon_T = \begin{pmatrix} T\cdot I_d & \underset{d\times d}{0} \\ \underset{d\times d}{0} & I_d \end{pmatrix}.$$

We can write the pseudo-regression in vector form as

$$\mathbb{W}^y \approx \tilde{\zeta}\left(\Upsilon_T \cdot \gamma_0\right) + \tilde{\nu},$$

where $\tilde{\nu} \perp \tilde{\zeta}$ and $\tilde{\nu} \sim N(0, \sigma_{0\cdot x}^2 I_K)$. The theorem below follows easily from the above approximate formulation. A rigorous proof is given in the appendix.

Theorem 1 *Let Assumptions 1 and 2 hold. Then under the fixed-K asymptotics where K is held fixed as $T \to \infty$, we have*

$$\Upsilon_T\left(\hat{\gamma}_{TAOLS} - \gamma_0\right) \Rightarrow (\tilde{\zeta}'\tilde{\zeta})^{-1}\tilde{\zeta}'\tilde{\nu}.$$

A direct implication of Theorem 1 is that

$$T(\hat{\beta}_{TAOLS} - \beta_0) \Rightarrow \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu, \tag{25}$$

$$\hat{\delta}_{TAOLS} - \delta_0 \Rightarrow \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \nu, \tag{26}$$

where $M_\xi = I_K - \xi\left(\xi'\xi\right)^{-1}\xi'$ and $M_\eta = I_K - \eta\left(\eta'\eta\right)^{-1}\eta'$. Conditional on (η, ξ), both limiting distributions are normal:

$$\sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu \overset{d}{=} N\left[0, \sigma_{0\cdot x}^2 \Omega_{xx}^{-1/2}(\eta' M_\xi \eta)^{-1}\Omega_{xx}^{-1/2}\right],$$

$$\sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi' M_\eta \xi\right)^{-1}\xi' M_\eta \nu \overset{d}{=} N\left[0, \sigma_{0\cdot x}^2 \Omega_{xx}^{-1/2}(\xi' M_\eta \xi)^{-1}\Omega_{xx}^{-1/2}\right].$$

Therefore, the unconditional limiting distributions are mixed normal. Furthermore, there is no second-order endogeneity bias in the TAOLS estimator. The TAOLS approach has effectively removed the two problems that plague the naive OLS estimator. The first problem, i.e., the asymptotic dependence between the partial sum processes of the regressor and the regression error is eliminated because we augment the original regression by the additional regressor Δx_t. The second problem, i.e., the second-order endogeneity bias, is eliminated because we transform the original data and run the regression in the space spanned by the basis functions. In general, both augmentation and transformation are necessary to achieve asymptotic mixed normality and asymptotic unbiasedness. However, for some special basis functions, augmentation is not necessary for the asymptotic mixed normality. See Section 4 for more detail.

The key result that drives the asymptotic unbiasedness of the TAOLS estimator is that

$$\Upsilon_T^{-1}\tilde{\mathbb{W}}'\mathbb{W}^{0\cdot x} = \sum_{i=1}^{K}(\Upsilon_T^{-1}\tilde{\mathbb{W}}_i)\mathbb{W}_i^{0\cdot x} \Rightarrow \tilde{\zeta}'\tilde{\nu}, \tag{27}$$

which is mixed normal with mean zero. Note that the corresponding term in the OLS estimator based on (13) is

$$\Upsilon_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{x}_t \cdot u_{0\cdot xt}$$

for $\tilde{x}_t = (x_t', \Delta x_t')'$. It is well known that this term has an additive nuisance bias in the limit, which necessitates the $\mathcal{M}$ correction in (11). In contrast, the limit in (27) does not have such an additive bias term. Like the "partial sum" transforms in Vogelsang and Wagner (2014), basis-function transforms help eliminate the additive bias. Both types of transforms resemble low-pass filters that reinforce low-frequency components while attenuating high-frequency components. However, there is an important difference. While basis function transforms make the error process asymptotically independent, "partial sum" transforms make it more persistent with a consequential adverse effect on the asymptotic efficiency. See Vogelsang and Wagner (2014) for more discussion on the efficiency of the IMOLS estimator as well as its finite-sample bias-reduction property in certain scenarios.

We note that the TAOLS estimator $\hat{\delta}$ of δ_0 is not consistent when K is fixed. The consistency of $\hat{\delta}$ can be restored if we consider a different limiting thought experiment where K increases with T but at a slower rate. See Phillips (2014) for details.

Proposition 2 *Let Assumptions 1 and 2 hold.*

(i) Under the fixed-K asymptotics, $T(\hat{\beta}_{TIV} - \beta_0) = T(\hat{\beta}_{TAOLS} - \beta_0) + o_p(1)$.

(ii) Let V_K *be a random variable with distribution* $MN\left[0, \sigma^2_{0\cdot x}\Omega_{xx}^{-1/2}(\eta' M_\xi \eta)^{-1}\Omega_{xx}^{-1/2}\right]$. *Assume that* $\{\phi_i(\cdot)\}_{i=1}^{\infty}$ *is a complete orthonormal system in*

$$L_0^2[0,1] = \left\{ f(\cdot) \in L^2[0,1] : \int_0^1 f(r)\, dr = 0 \right\}.$$

Then as $K \to \infty$,

$$V_K \Rightarrow MN\left[0, \sigma^2_{0\cdot x}\Omega_{xx}^{-1/2}\left(\int_0^1 \tilde{W}_x(r)\tilde{W}_x(r)' dr\right)^{-1}\Omega_{xx}^{-1/2}\right]$$

where $\tilde{W}_x(r) = W_x(r) - \int_0^1 W_x(s)\, ds$ *is the demeaned version of* $W_x(r)$.

Given the asymptotic equivalence in Proposition 2, our fixed-K asymptotic theory applies to the TIV estimator. This can be regarded as a by-product of our paper. The fixed-K asymptotic theory for the TIV estimator was established by Phillips and Liao (2014, Lemma 5.1), but they considered only the case with a scalar regressor and did not pursue the limit t approximation theory established in this paper.

The conditional variance in Proposition 2(ii) is the semiparametric efficiency bound in the sense of Phillips (1991b). Here we do not aim at achieving the bound *per se*. Instead, our goal is to come up with a more accurate approximation for the given K value in a finite sample situation. Proposition 2(ii) indicates that the TAOLS estimator could become more efficient for a larger K and ultimately reach the semiparametric efficiency bound under the increasing-K asymptotics. So, from this alternative asymptotic point of view, there is no loss of efficiency in our TAOLS approach.

The asymptotics in Proposition 2(ii) is obtained for a fixed K as $T \to \infty$ and then letting $K \to \infty$. This is a type of sequential asymptotics. The sequential asymptotics provides a smooth transition from our fixed-K asymptotics to the increasing-K asymptotics in Phillips (2005, 2014).

We note that for the TIV estimator Phillips (2014) considers only the increasing-K asymptotics under which T and K go to infinity and $K/T \to 0$. A careful inspection of his proof shows that it applies to the TAOLS estimator as well. Thus, in effect, Phillips (2014) has also established the increasing-K asymptotics for the TAOLS estimator. More specifically, assume that $\{\phi_i(\cdot)\}_{i=1}^{\infty}$ is a complete orthonormal system in

$$L_0^2[0,1] := \left\{ f(\cdot) \in L^2[0,1] : \int_0^1 f(r)\, dr = 0 \right\}.$$

Then under the linear process assumption and rate condition given in his equations (L) and (R), we have

$$T(\hat{\beta}_{TAOLS}-\beta_0) \Rightarrow MN\left[0, \sigma^2_{0\cdot x}\Omega_{xx}^{-1/2}\left(\int_0^1 \tilde{W}_x(r)\tilde{W}_x(r)'dr\right)^{-1}\Omega_{xx}^{-1/2}\right], \tag{28}$$

where $\tilde{W}_x(r) = W_x(r) - \int_0^1 W_x(s)\,ds$ is the demeaned version of $W_x(r)$. The above result is slightly different from what is given in the main theorem of Phillips (2014, page 213). The difference arises because we require the basis functions to integrate to zero in order to accommodate an intercept in the cointegration model. The cointegration model considered in Phillips (2014) has no intercept, and so the basis functions do not have to integrate to zero. This difference is innocuous, and the proof in Phillips (2014) goes through with only minor modifications.

The asymptotic mixed normality and unbiasedness of the TAOLS estimator facilitate hypothesis testing. Suppose that we are interested in testing

$$H_0 : R\beta_0 = r \text{ vs. } H_1 : \; R\beta_0 \neq r, \tag{29}$$

where R is a $p \times d$ matrix. If $\sigma^2_{0\cdot x}$ is known, then we would construct the following Wald statistic:

$$\tilde{F}(\hat{\beta}_{TAOLS}) = \frac{1}{\sigma^2_{0\cdot x}}(R\hat{\beta}_{TAOLS}-r)'\left[R(\mathbb{W}^{x\prime}M_{\Delta x}\mathbb{W}^{x})^{-1}R'\right]^{-1}(R\hat{\beta}_{TAOLS}-r)/p.$$

When $p = 1$, we would construct the following t statistic:

$$\tilde{t}(\hat{\beta}_{TAOLS}) = \frac{R\hat{\beta}_{TAOLS}-r}{\sqrt{\sigma^2_{0\cdot x}R(\mathbb{W}^{x\prime}M_{\Delta x}\mathbb{W}^{x})^{-1}R'}}.$$

Under the null hypothesis in (29), we can invoke Theorem 1 to obtain

$$\tilde{F}(\hat{\beta}_{TAOLS}) \Rightarrow Q'[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}']^{-1}Q/p, \tag{30}$$

where

$$\tilde{R} = R\Omega_{xx}^{-1/2} \text{ and } Q = \tilde{R}(\eta' M_\xi \eta)^{-1}\eta' M_\xi \nu. \tag{31}$$

By construction, Q follows the mixed normal distribution $MN\left[0, \tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'\right]$. Conditional on $\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'$,

$$Q'[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}']^{-1}Q/p \sim \chi^2_p/p.$$

The conditional distribution does not depend on the conditioning variable $\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'$. So χ^2_p/p is also the unconditional distribution. That is, the infeasible test statistic $\tilde{F}(\hat{\beta}_{TAOLS})$ converges in distribution to χ^2_p/p. Similarly, $\tilde{t}(\hat{\beta}_{TAOLS})$ converges in distribution to the standard normal distribution.

The presence of the unknown long run variance $\sigma^2_{0\cdot x}$ in $\tilde{F}(\hat{\beta}_{TAOLS})$ and $\tilde{t}(\hat{\beta}_{TAOLS})$ hinders their practical application. In practice, we have to estimate $\sigma^2_{0\cdot x}$ in order to construct the test statistics. Given that $\sigma^2_{0\cdot x}$ is the approximate variance of the error term in the TAOLS regression, it is natural to estimate $\sigma^2_{0\cdot x}$ by

$$\hat{\sigma}^2_{0\cdot x} = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\mathbb{W}}_i^{0\cdot x}\right)^2 = \frac{1}{K}\mathbb{W}^{0\cdot x\prime}\left[I_K - \tilde{\mathbb{W}}(\tilde{\mathbb{W}}'\tilde{\mathbb{W}})^{-1}\tilde{\mathbb{W}}'\right]\mathbb{W}^{0\cdot x},$$

where $\hat{\mathbb{W}}_i^{0\cdot x} = \mathbb{W}_i^y - \mathbb{W}_i^{x\prime}\hat{\beta}_{TAOLS} - \mathbb{W}_i^{\Delta x\prime}\hat{\delta}_{TAOLS}$. With the estimator $\hat{\sigma}_{0\cdot x}^2$, we can construct the feasible $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$ as follows:

$$F(\hat{\beta}_{TAOLS}) = \frac{1}{\hat{\sigma}_{0\cdot x}^2}(R\hat{\beta}_{TAOLS} - r)' \left[R(\mathbb{W}^{x\prime} M_{\Delta x}\mathbb{W}^x)^{-1}R'\right]^{-1}(R\hat{\beta}_{TAOLS} - r)/p, \tag{32}$$

$$t(\hat{\beta}_{TAOLS}) = \frac{R\hat{\beta}_{TAOLS} - r}{\sqrt{\hat{\sigma}_{0\cdot x}^2 R(\mathbb{W}^{x\prime} M_{\Delta x}\mathbb{W}^x)^{-1}R'}}.$$

The theorem below establishes the limiting null distributions of $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$ under the fixed-K asymptotics.

Theorem 3 *Let Assumptions 1 and 2 hold. Under the fixed-K asymptotics, we have*

$$F(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d} \cdot F_{p,K-2d} \text{ and}$$
$$t(\hat{\beta}_{TAOLS}) \Rightarrow \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d},$$

where $F_{p,K-2d}$ is the F distribution with degrees of freedom p and $K-2d$, and t_{K-2d} is the t distribution with degrees of freedom $K-2d$.

Theorem 3 shows that both $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$ are asymptotically pivotal and have standard limiting distributions. From an asymptotic point of view, the TA regression is equivalent to a classical linear normal regression (CLNR), and so the F and t limit theory is not surprising. The standard F and t limit distributions we obtain here are in sharp contrast with the nonstandard limiting distributions in Vogelsang and Wagner (2014). A great advantage of our approximations is that the critical values can be obtained from statistical tables and software packages. There is no need to simulate nonstandard critical values.

Our asymptotic distributions are also in sharp contrast to the chi-squared (χ_p^2/p) and standard normal distributions. The latter distributions are the weak limits for the infeasible test statistics. In fact, under the increasing-K asymptotics as developed in Phillips (2014), the latter distributions are also the weak limits of the feasible statistics $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$. So the increasing-K asymptotics effectively assume that $\sigma_{0\cdot x}^2$ is known in large samples, and hence completely ignore the estimation uncertainty in $\hat{\sigma}_{0\cdot x}^2$.

To compare the critical values from the fixed-K approximation with those from the increasing-K approximation, we consider the Wald-type test as an example. Let $F_{p,K-2d}^{\alpha}$ and χ_p^{α} be the $(1-\alpha)$ quantiles from the standard $F_{p,K-2d}$ and χ_p^2 distributions, respectively. Then we can use the modified F critical value $K/(K-2d)F_{p,K-2d}^{\alpha}$ to carry out our F test. This critical value is larger than the scaled chi-squared critical value χ_p^{α}/p for two reasons. First, $F_{p,K-2d}^{\alpha} > \chi_p^{\alpha}/p$, because the F distribution $F_{p,K-2d}$ has a random denominator, unlike the corresponding chi-squared distribution. Second, the multiplicative factor $K/(K-2d)$ is greater than 1. The difference between the two critical values depends on the value of K. It can be quite large when K is small. However, as K increases, $K/(K-2d)F_{p,K-2d}^{\alpha}$ approaches χ_p^{α}/p. There is a smooth transition from a fixed-K critical value to the corresponding increasing-K critical value. So, the critical value $K/(K-2d)F_{p,K-2d}^{\alpha}$ is asymptotically valid regardless of whether K is held fixed or allowed to grow with the sample size. In this sense, $K/(K-2d)F_{p,K-2d}^{\alpha}$ is a robust critical value.

To investigate the power of the F and t tests, we consider the local alternative hypothesis

$$H_{1T} : R\beta_0 = r + \theta/T \text{ for some } \theta \in \mathbb{R}^p.$$

The following theorem establishes the limiting distributions of $F(\hat{\beta}_{TAOLS})$ and $t(\hat{\beta}_{TAOLS})$ under H_{1T}.

Theorem 4 *Let Assumptions 1 and 2 hold. Under H_{1T}, we have*

$$F(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d} F_{p,K-2d}(\|\lambda\|^2)$$
$$t(\hat{\beta}_{TAOLS}) \Rightarrow \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d}(\lambda) \text{ for } p = 1$$

as $T \to \infty$ for a fixed K, where

$$\lambda = \frac{\left[R\Omega_{xx}^{-1/2} (\eta' M_\xi \eta)^{-1} \Omega_{xx}^{-1/2} R'\right]^{-1/2} \theta}{\sigma_{0\cdot x}},$$

and $F_{p,K-2d}(\cdot)$ and $t_{K-2d}(\cdot)$ are noncentral F and t distributions with noncentrality parameters $\|\lambda\|^2$ and λ, respectively.

Since λ is random, the asymptotic distributions are mixed noncentral F and t distributions. The mixed distributions are analogous to the mixed chi-squared or normal distribution we would get in the conventional FMOLS framework. More broadly, asymptotic mixed noncentral distributions are typical for experiments that have the local asymptotic mixed normality property.

Under the null hypothesis, Theorem 3 shows that the basis functions have no effect on the asymptotic distributions. Under the local alternative, Theorem 4 shows that the effect of the basis functions on the asymptotic distributions is manifested through the noncentrality parameter λ only. Let $\boldsymbol{\phi}(x) = (\phi_1(x), \ldots, \phi_K(x))'$ be the vector of basis functions and A be any orthogonal matrix, then it is easy to see that λ will not change if $A\boldsymbol{\phi}(x)$ is used as the vector of basis functions instead. A direct implication is that the power of the F or t test is invariant to rotations of the basis functions.

For a given value of K, we may choose the basis functions to maximize the local asymptotic power, say $P[F_{p,K-2d}(\|\lambda\|^2) > F^{\alpha}_{p,K-2d}]$. This is not an easy task, as, in general, ξ and η are not independent of each other. The optimal choice may also depend on $R\Omega_{xx}^{-1/2}$ and the direction of the local departure characterized by θ. We leave this to future research.

3.2 Model with a linear trend

In this subsection, we consider a more general version of (1) by including a linear time trend in the cointegration model. The model is now given by

$$y_t = x_t'\beta_0 + \mu_0 t + u_{0t}, \tag{33}$$
$$x_t = x_{t-1} + u_{xt}.$$

Define $\mathbb{W}_i^{tr} = T^{-1/2}\sum_{t=1}^{T} \phi_i(t/T)t$ for $i = 1, \ldots, K$ and $\mathbb{W}^{tr} = (\mathbb{W}_1^{tr}, \ldots, \mathbb{W}_K^{tr})'$. Then the transformed regression in (19) is naturally generalized to

$$\mathbb{W}_i^y = \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{tr}\mu_0 + \mathbb{W}_i^{0\cdot x} \text{ for } i = 1, \ldots, K. \tag{34}$$

As we discussed earlier, an intercept can be included in (33) and (34), but our approach is asymptotically invariant to location shifts. The TAOLS estimator for β_0, δ_0, and μ_0 is now given by

$$(\hat{\beta}'_{TAOLS}, \hat{\delta}'_{TAOLS}, \hat{\mu}'_{TAOLS})' = (\tilde{\mathbb{W}}'_{tr}\tilde{\mathbb{W}}_{tr})^{-1}\tilde{\mathbb{W}}'_{tr}\mathbb{W}^{y}, \tag{35}$$

where $\tilde{\mathbb{W}}_{tr} = \left(\mathbb{W}^{x},\ \mathbb{W}^{\Delta x}, \mathbb{W}^{tr}\right)$.

Let

$$\begin{aligned} \hat{\mathbb{W}}^{0\cdot x}_{i,tr} &= \mathbb{W}^{y}_{i} - \mathbb{W}^{x\prime}_{i}\hat{\beta}_{TAOLS} - \mathbb{W}^{\Delta x\prime}_{i}\hat{\delta}_{TAOLS} - \mathbb{W}^{tr}_{i}\hat{\mu}_{TAOLS}, \\ (\hat{\sigma}^{tr}_{0\cdot x})^2 &= K^{-1}\sum_{i=1}^{K}(\hat{\mathbb{W}}^{0\cdot x}_{i,tr})^2. \end{aligned}$$

Then we can construct the Wald statistic and t statistic as follows:

$$\begin{aligned} F_{tr}(\hat{\beta}_{TAOLS}) &= \frac{1}{(\hat{\sigma}^{tr}_{0\cdot x})^2}(R\hat{\beta}_{TAOLS} - r)\left[R(\mathbb{W}^{x\prime}M_{\Delta x,tr}\mathbb{W}^{x})^{-1}R'\right]^{-1}(R\hat{\beta}_{TAOLS} - r)/p, \\ t_{tr}(\hat{\beta}_{TAOLS}) &= \frac{R\hat{\beta}_{TAOLS} - r}{\sqrt{(\hat{\sigma}^{tr}_{0\cdot x})^2 R(\mathbb{W}^{x\prime}M_{\Delta x,tr}\mathbb{W}^{x})^{-1}R'}}, \end{aligned}$$

where $M_{\Delta x,tr} = I_K - \mathbb{W}_{\Delta x,tr}\left(\mathbb{W}'_{\Delta x,tr}\mathbb{W}_{\Delta x,tr}\right)^{-1}\mathbb{W}'_{\Delta x,tr}$ and $\mathbb{W}_{\Delta x,tr} = \left(\mathbb{W}^{\Delta x}, \mathbb{W}^{tr}\right)$.

Theorem 5 *Let Assumptions 1 and 2 hold. Assume that* $a := \left(\int_0^1 \phi_1(r)\, rdr, \cdots, \int_0^1 \phi_K(r)\, rdr\right)' \neq 0$. *Under the fixed-K asymptotics, we have (i)*

$$\Upsilon_{T,tr}\begin{pmatrix} \hat{\beta}_{TAOLS} - \beta_0 \\ \hat{\delta}_{TAOLS} - \delta_0 \\ \hat{\mu}_{TAOLS} - \mu_0 \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta' M_{\xi,a}\eta\right)^{-1}\eta' M_{\xi,a}\nu \\ \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi' M_{\eta,a}\xi\right)^{-1}\xi' M_{\eta,a}\nu \\ \sigma_{0\cdot x}\left(a' M_{\eta,\xi}a\right)^{-1}a' M_{\eta,\xi}\nu \end{pmatrix}, \tag{36}$$

where

$$\Upsilon_{T,tr} = \begin{pmatrix} \Upsilon_T & 0 \\ 0 & T^{3/2} \end{pmatrix}$$

and M_{υ} is the projection matrix projecting onto the orthogonal complement of the column space of υ.

(ii) Under the null hypothesis $H_0 : R\beta_0 = r$, we have

$$F_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d-1}F_{p,K-2d-1} \text{ and } t_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \sqrt{\frac{K}{K-2d-1}}t_{K-2d-1}. \tag{37}$$

(iii) Under the local alternative hypothesis $H_{1T} : R\beta_0 = r + \theta/T$, we have

$$F_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d-1}F_{p,K-2d-1}(\|\lambda\|^2) \tag{38}$$

$$\text{and } t_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \sqrt{\frac{K}{K-2d-1}}t_{K-2d-1}(\lambda), \tag{39}$$

where

$$\lambda = \frac{\left[R\Omega_{xx}^{-1/2}\left(\eta' M_{\xi,a}\eta\right)^{-1}\Omega_{xx}^{-1/2}R'\right]^{-1/2}\theta}{\sigma_{0\cdot x}}.$$

Theorems 5(ii) and (iii) are entirely analogous to Theorems 3 and 4. The effect of having an additional trend regressor $\mathbb{W}_i^{tr}$ is reflected by the adjustment in the multiplicative correction factor and the degrees of freedom in the limiting F and t distributions.

Again, the asymptotic F and t limit theory resembles the standard theory in the CLNR with K iid observations. The multiplicative correction is a type of degrees-of-freedom correction. Had we followed the standard practice in the CLNR and defined

$$(\hat{\sigma}_{0\cdot x}^{tr})^2 = \frac{1}{K-2d-1}\sum_{i=1}^{K}(\hat{\mathbb{W}}_{i,tr}^{0\cdot x})^2, \tag{40}$$

we would not have to make the multiplicative correction. That is, the (scaled) Wald statistic would be asymptotically F distributed, and the t statistic would be asymptotically t distributed.

Observing that we compute the standard error of the TAOLS estimator in the same way we would if the errors in the transformed regression are homoskedastic, which does hold in large samples, our Wald statistic $F_{tr}(\hat{\beta}_{TAOLS})$ with (40) as the error variance estimator is numerically identical to the F statistic based on the residual sum of squares under the restricted and unrestricted models. So, we can obtain $F_{tr}(\hat{\beta}_{TAOLS})$ (and $t_{tr}(\hat{\beta}_{TAOLS})$) from the output of any simple and very basic regression program as long as it works at least for the CLNR with homoskedastic errors. The only step that we have to take is to get the data into the transformed form. It should be noted, however, that we do not include the intercept in the transformed and augmented regression when the basis functions satisfy $\int_0^1 \phi_i(x)\,dx = 0$.

As a by-product, we can perform a test of endogeneity, that is, the test of whether $\delta_0 = 0$, in exactly the same way we would if the transformed regression is a CLNR. This can be justified asymptotically using the same argument as in the proof of Theorem 5. We note that the test will be inconsistent for a fixed K, but our focus here is on obtaining more accurate approximations. The fixed-K asymptotics do not require that we fix the value of K in finite samples. In fact, in empirical applications the sample size T is usually given beforehand, and the value of K needs to be determined using *a priori* information and/or information obtained from the data. While the selected value of K may be relatively large for large T, it is also true that it is a finite value for any given sample. Plugging this finite value into the fixed-K asymptotic distribution gives us a practical way to use the fixed-K approximation even when K is data-driven. As we have already shown, the fixed-K critical value so obtained is asymptotically valid under the increasing-K asymptotics.

If we have the polynomial trends $(t, t^2, \ldots, t^g)$ for any integer g instead of a linear trend, then the proof of Theorem 5 can be invoked to establish the asymptotic distributions under the null and the local alternative. For example, under the null $H_0: R\beta_0 = r$, we can show that when $K > 2d + \tilde{g}$,

$$F_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d-g}F_{p,K-2d-\tilde{g}} \text{ and } t_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \sqrt{\frac{K}{K-2d-g}}t_{K-2d-\tilde{g}}, \tag{41}$$

where

$$\tilde{g} = rank\left(\int \left(\phi_1(r), \cdots, \phi_K(r)\right)' \left(r, r^2, \cdots, r^g\right) dr\right) \tag{42}$$

and $2d + \tilde{g}$ is now the effective number of parameters to be estimated. The finite-sample analogue of $\tilde{g}$ is just the rank of $\mathbb{W}^{tr}$, the transformed trend matrix. If $clsp\{r, \cdots, r^g\} \cap [clsp\{\phi_1(r), \cdots, \phi_K(r)\}]^{\perp} = \{0\}$, where '$clsp$' stands for the closed linear span, then $\tilde{g} = g$. If some nonzero linear combination of the polynomial trends belongs to $clsp\left\{\phi_{K+1}(r), \phi_{K+2}(r), \cdots\right\}$, then $\tilde{g} < g$. For the Fourier basis functions, $clsp\{\phi_1(r), \cdots, \phi_K(r)\}$ is a lower-frequency space, as the energy of each non-zero function in this space concentrates only on lower frequencies. Correspondingly, its orthogonal complement $[clsp\{\phi_1(r), \cdots, \phi_K(r)\}]^{\perp}$ is a higher-frequency space. Given that there does not exist any nontrivial linear combination of the polynomial trends whose energy concentrates only on the higher frequencies, we have $clsp\{r, \cdots, r^g\} \cap [clsp\{\phi_1(r), \cdots, \phi_K(r)\}]^{\perp} = \{0\}$, and so $\tilde{g} = g$.

The above comments apply to any set of trend functions. For more general trend functions, we define $\tilde{g}$ in the same manner as in (42) but with $\left(r, r^2, \cdots, r^g\right)$ replaced by the general trend functions. In some statistical packages such as STATA, we do not even need to compute $\tilde{g}$ theoretically. When $\tilde{g} < g$ and the sample size is large enough, STATA will notice a multicollinearity problem and retain only $\tilde{g}$ transformed trend variables to avoid the multicollinearity. The asymptotic F test and t test can then be performed based on the new transformed and augmented regression.

It is important to point out that under the fixed-K asymptotics the exact forms of the trend functions and hence their transforms have to be known in order to ensure the consistency of the TAOLS estimator $\hat{\beta}_{TAOLS}$. If the trend functions are misspecified, then the TAOLS estimator is in general inconsistent. However, if K is large and the trend functions are smooth enough that they can be well approximated by a linear combination of a sufficiently large subset of the basis functions, we can include this subset of basis functions in the TA regression. This will help control for the unknown trend functions and reduce the asymptotic bias of the TAOLS estimator. To eliminate the asymptotic bias altogether, we have to let K grow with the sample size T at some rate, and we are no longer in the domain of the fixed-K asymptotics but instead move into the domain of the increasing-K asymptotics. So, if K is held fixed literally, then the fixed-K asymptotics is not robust to trend misspecification. This is in contrast with the increasing-K asymptotics under which the TAOLS estimator can still be made consistent, even if the exact forms of the trend functions are not known. However, as we discussed before, we do not have to fix the value of K in order to use the fixed-K asymptotic approximations. Even if K grows with the sample size T, we can still use the fixed-K critical values (i.e., F and t critical values), as they remain valid under the increasing-K asymptotics.

3.3 The form of basis functions

We consider the following two sets of basis functions on $L^2[0,1]$:

Fourier basis functions

$$\left\{\phi_{2j-1}(r) = \sqrt{2}\cos(2j\pi r),\ \phi_{2j} = \sqrt{2}\sin(2j\pi r), j = 1, \ldots, K/2\right\}; \tag{43}$$

Cosine basis functions

$$\left\{\phi_j(r) = \sqrt{2}\cos(j\pi r),\ j = 1, \ldots, K\right\}. \tag{44}$$

Both sets contain orthonormal basis and satisfy Assumption 1. If we let $K \to \infty$, then the orthonormal basis functions in both sets are complete. Fourier basis functions are the standard and commonly used basis functions. Cosine basis functions are the eigenfunctions of the covariance kernel of the demeaned Brownian motion.

Another commonly used set of trigonometric basis functions consists of the sine basis functions $\{\sqrt{2}\sin(j\pi r)\}_{j=1}^{K}$ and $\{\sqrt{2}\sin((j-1/2)\pi r)\}_{j=1}^{K}$, which are the eigenfunctions of the covariance kernel of the standard Brownian bridge and Brownian motion, respectively. However, they cannot be used directly, as they do not satisfy Assumption 1(ii).

From a theoretical point of view, any orthonormal basis can be used. For example, one may use the Legendre polynomials (shifted and renormalized):

$$\left\{\phi_j(r) = \sqrt{2j+1}P_j(2r-1), \ j = 1, \ldots, K\right\}, \tag{45}$$

where

$$P_j(s) = \frac{1}{2^j j!}\frac{d^j}{ds^j}\left(s^2-1\right)^j \text{ for } s \in [-1,1]$$

are the standard Legendre polynomials. Our simulation study not reported here shows that Legendre polynomials do not deliver an accurate F test or t test in finite samples when high-order polynomials are used. The main reason is that while these polynomials are orthonormal in $L^2[0,1]$, i.e., $\int_0^1 \phi_i(r)\phi_j(r)\,dr = 1\{i=j\}$, the discrete version of the integral, namely, $T^{-1}\sum_{t=1}^T \phi(i/T)\phi(j/T)$, may not be close to $1\{i=j\}$, especially when i and j are large. So even if $u_{0\cdot x,t}$ is iid $N(0,\sigma^2)$, the transformed error $\mathbb{W}_i^{0\cdot x}$ may be far from being iid $N(0,\sigma^2)$ in finite samples. This can lead to substantial finite-sample size distortion. In contrast, for the two sets of trigonometric basis functions given earlier, we can show that because of cancellations,

$$\left|T^{-1}\sum_{t=1}^T \phi(i/T)\phi(j/T) - 1\{i=j\}\right| \leq 2/T$$

for all integers $i, j \in [1,T)$. In fact, for Fourier basis functions, $T^{-1}\sum_{t=1}^T \phi(i/T)\phi(j/T)$ is exactly equal to $1\{i=j\}$. For this reason, we recommend using the trigonometric basis functions, and hereinafter we do not consider other basis functions such as Legendre polynomials.

3.4 The number of basis functions

In principle, we can use any finite number of orthonormal basis functions in our fixed-K framework. However, Proposition 2 indicates that a larger K leads to a more efficient estimator. On the other hand, when K is too large, the TAOLS estimator will suffer from the asymptotic bias that is not captured by the fixed-K asymptotics. For example, if we set K equal to the sample size, which is the upper bound for K, the TAOLS estimator will be the same as the augmented OLS estimator, which suffers from second-order asymptotic bias. Thus there is an opportunity to select K to trade off the variance effect with the bias effect.

A direct approach to a data-driven choice of K is to first develop a high-order expansion of $\hat{\beta}_{TAOLS}$ from which we obtain the approximate mean squared error (AMSE) of $\hat{\beta}_{TAOLS}$ and then select K to minimize the AMSE of $\hat{\beta}_{TAOLS}$. For hypothesis testing, a direct approach is to choose K to minimize the Type II error of our proposed F test or t test subject to a control of the Type I error. The direct approaches are ambitious. Phillips (2014) discusses some of the technical challenges behind the direct approaches. We leave them for future research.

An indirect approach that appears to work reasonably well is based on the bias and variance of the LRV estimator. Following a large literature on LRV estimation, Phillips (2005b) proposes selecting K by minimizing the AMSE of $\hat{\Omega}$ defined in (8). In the present setting, we have

$$\hat{\Omega} = \frac{1}{K}\sum_{i=1}^K \left(\hat{\mathbb{W}}_i^u\right)\left(\hat{\mathbb{W}}_i^u\right)' \text{ for } \hat{\mathbb{W}}_i^u = \frac{1}{\sqrt{T}}\sum_{t=1}^T \hat{u}_t\phi_i(\frac{t}{T}),$$

where either $\hat{u}_t = (y_t - x_t \hat{\beta}_{OLS}, \Delta x_t')'$ or $\hat{u}_t = (y_t - x_t \hat{\beta}_{OLS} - \hat{\mu}_{OLS} t, \Delta x_t')'$, depending on whether a linear trend is present or not. For the Fourier basis functions and cosine basis functions, the AMSE-optimal K^* for estimating $\hat{\Omega}$ is given by

$$K^*_{MSE} = \left\lceil \left(\frac{tr(I_{m^2} + \mathbb{K}_{mm})(\Omega \otimes \Omega)}{4vec(B)' \, vec(B)} \right)^{1/5} T^{4/5} \right\rceil$$

$$\text{for } B = -\frac{\pi^2}{6} \sum_{h=-\infty}^{\infty} h^2 \Gamma_u(h), \ \Gamma_u(h) = E u_t u_{t-h}', \tag{46}$$

where $\mathbb{K}_{mm}$ is the $m^2 \times m^2$ commutation matrix and I_{m^2} is the $m^2 \times m^2$ identity matrix. It is important to point out that the above formula is based on the AMSE of the LRV estimator $\hat{\Omega}$, which is related to the TAOLS estimator but is essentially a different problem. Therefore, the above rule of selecting K should be regarded as only a rule of thumb.

Recall that K has to be large enough to ensure that the regressors in the TA regression are not perfectly multicollinear. In the absence of a trend, it is necessary to have $K \geq 2d$. For one of his tests, Bierens (1997) recommends the rule-of-thumb value $K = 2d$. In our setting, the limiting distribution of the Wald statistic is the F distribution with the denominator degrees of freedom $K - 2d$. For this F distribution to have a finite mean, we require $K - 2d \geq 3$, i.e., $K \geq 2d + 3$. So in finite samples it is reasonable to set K equal to $K^*_{MSE,c}$ with

$$K^*_{MSE,c} = \max(2d + 3, K^*_{MSE}). \tag{47}$$

When a linear trend is included, we make an obvious adjustment and set K equal to the following $K^*_{MSE,c}$:

$$K^*_{MSE,c} = \max(2d + 4, K^*_{MSE}). \tag{48}$$

There is another reason to avoid a large K. Cointegration is fundamentally a long run relationship. To estimate the cointegrating vector, we should employ a regression that uses only the long run variation of the underlying variables. The short run variation can help only when the short run relationship coincides with the long run relationship. If the two types of relationships differ from each other, then going beyond a reasonable value of K runs the risk of being struck by short run contaminations. A trade-off between the asymptotic efficiency and robustness with respect to short run contaminations leads us to consider a moderate K value.

When the Fourier basis functions (43) are used, the transformed data consist of the real and imaginary parts of the discrete Fourier transforms of the original data. In this case, a useful rule-of-thumb choice is suggested by Müller (2014) and Müller and Watson (2013). These papers propose selecting a K value to reflect business-cycle frequencies or below. For example, with $T = 64$ years of post-World-War-II macro data, the choice of $K = 16$ captures the long run movements of macro data with periodicity higher than the commonly accepted business-cycle period of $T/(K/2) = 8$ years. Most recently, after extensive simulations, Lazarus, Lewis, Stock and Watson (2016) suggest choosing $K = 8$ for HAR inference. To follow the fixed-K spirit in the strictest sense, we will consider both $K = 8$ and $K = 16$ in our simulations.

4 Cointegration Analysis with Shifted Cosine Bases

4.1 Model without time trend

We go back to the model without an intercept and time trend, i.e., the model in (19), but we drop the augmented term in (20) and consider the following equation

$$\mathbb{W}_i^y = \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^0 \tag{49}$$

where by definition

$$\mathbb{W}_i^0 = \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{0t}\phi_i\left(\frac{t}{T}\right).$$

Define the transformed OLS (TOLS) estimator $\hat{\beta}_{TOLS}$ to be

$$\hat{\beta}_{TOLS} = \left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1}\mathbb{W}^{x\prime}\mathbb{W}^y. \tag{50}$$

In general, dropping $\mathbb{W}_i^{\Delta x}$ in (20) will lead to an omitted variable bias unless the correlation between $\mathbb{W}^{\Delta x}$ and $\mathbb{W}^x$ is zero. The zero correlation is ensured by the following assumption.

Assumption 3 *The basis functions satisfy $\int_0^1 \Psi_i(r)\phi_j(r)\,dr = 0$ with $\Psi_i(r) = \int_0^r \phi_i(s)\,ds$ for $i, j = 1, ..., K$.*

Recall that

$$\begin{aligned}\frac{\mathbb{W}_i^x}{T} &\Rightarrow \Omega_{xx}^{1/2}\int_0^1 \phi_i(r)W_x(r)dr = -\Omega_{xx}^{1/2}\int_0^1 \Psi_i(r)dW_x(r),\\ \mathbb{W}_j^0 &\Rightarrow \sigma_{0\cdot x}\int_0^1 \phi_j(r)\,dw_0(r) + \sigma_{0x}\Omega_{xx}^{-1/2}\int_0^1 \phi_j(r)dW_x(r),\end{aligned} \tag{51}$$

for $i, j = 1, ..., K$ where $w_0(r)$ and $W_x(r)$ are independent Brownian motion processes. The asymptotic distribution of $(\mathbb{W}_i^x/T, \mathbb{W}_j^0)$ is jointly normal with covariance

$$\operatorname{cov}\left(\int_0^1 \phi_i(r)B_x(r)dr, \int_0^1 \phi_j(r)dB_0(r)dr\right) = -\Omega_{xx}^{1/2}\left(\int_0^1 \Psi_i(r)\phi_j(r)\,dr\right)\Omega_{xx}^{-1/2}\sigma_{x0}.$$

Thus, $T^{-1}\mathbb{W}_i^x$ and $\mathbb{W}_j^{\Delta x}$ are asymptotically independent if the basis functions satisfy Assumption 3.

Lemma 6 *The cosine functions*

$$\phi_j^c(r) = \sqrt{2}\cos(2j\pi r) \ \ \textit{for } j = 1, ..., K \tag{52}$$

satisfy Assumptions 1 and 3.

The lemma not only shows that Assumption 3 can hold but also gives the set of simple and commonly-used cosine functions as an example. Although there may be other functions that satisfy Assumption 3, we have the cosine functions in mind when developing the asymptotic results in this section. We are not aware of other commonly-used basis functions that also satisfy Assumption 3.

Theorem 7 *Consider the model in (19). Let Assumptions 1–3 hold. Under the fixed-K asymptotics, we have*

$$T(\hat{\beta}_{TOLS} - \beta_0) \Rightarrow MN\left(0, \sigma_0^2 \Omega_{xx}^{-1/2} (\eta'\eta)^{-1} \Omega_{xx}^{-1/2}\right).$$

It is interesting to see that the transformed OLS estimator is asymptotically unbiased and mixed normal. To some extent, the use of the special basis functions such as the cosine functions kills two birds with one stone. There is no need to augment the original regression in order to achieve the asymptotic mixed normality.

Given the mixed normality of the limiting distribution, it is reasonable to make inference based on $\hat{\beta}_{TOLS}$. The Wald statistic and t statistic are

$$F_c(\hat{\beta}_{TOLS}) = \frac{1}{(\hat{\sigma}_0^c)^2}(R\hat{\beta}_{TOLS} - r)' \left[R\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1} R'\right]^{-1} (R\hat{\beta}_{TOLS} - r)/p, \tag{53}$$

$$t_c(\hat{\beta}_{TOLS}) = \frac{R\hat{\beta}_{TOLS} - r}{\hat{\sigma}_0^c \sqrt{R\left(\mathbb{W}^{x\prime}\mathbb{W}^x\right)^{-1} R'}}, \tag{54}$$

where

$$(\hat{\sigma}_0^c)^2 = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\mathbb{W}}_{ci}^0\right)^2 \text{ where } \hat{\mathbb{W}}_{ci}^0 = \mathbb{W}_i^y - \mathbb{W}_i^{x\prime}\hat{\beta}_{TOLS}.$$

Following a proof similar to that of Theorem 3, we can show that

$$F_c(\hat{\beta}_{TOLS}) \Rightarrow \frac{K}{K-d} \cdot F_{p,K-d} \text{ and } t_c(\hat{\beta}_{TOLS}) \Rightarrow \sqrt{\frac{K}{K-d}} \cdot t(K-d).$$

The above results are clearly analogous to the well-known results in a CLNR with K iid observations and d regressors.

4.2 Model with a linear trend

We consider the cointegration system with a linear trend as given in (33). Dropping the regressors $\mathbb{W}_i^{\Delta x}$ and $\mathbb{W}_i^{tr}$ in (34), we obtain

$$\mathbb{W}_i^y = \mathbb{W}_i^{x\prime}\beta_0 + \left(\mathbb{W}_i^{tr}\mu_0 + \mathbb{W}_i^0\right) \text{ for } i = 1, ..., K \tag{55}$$

where $\mathbb{W}_i^{tr}\mu_0 + \mathbb{W}_i^0 = \mathbb{W}_i^{tr}\mu_0 + \mathbb{W}_i^{\Delta x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x}$ is the composite error. In general, the transformed OLS estimator obtained by regressing $\mathbb{W}_i^y$ on $\mathbb{W}_i^x$ is not consistent even if cosine transforms are used. The reason is that the composite error is not mean zero and is correlated with the included regressor. In fact,

$$\begin{aligned}\mathbb{W}_i^{tr} &= \frac{\sqrt{2}}{\sqrt{T}}\sum_{t=1}^{T} t\cos(\frac{2\pi it}{T}) = T\sqrt{2T}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{t}{T}\cos(\frac{2\pi it}{T})\right] \\ &= T\sqrt{2T}\left[\int_0^1 r\cos\left(2\pi ir\right)dr + O\left(\frac{1}{T}\right)\right] = O(\sqrt{T})\end{aligned} \tag{56}$$

using $\int_0^1 r\cos\left(2\pi ir\right)dr = 0$. So the composite error grows with the sample size at the rate of $\sqrt{T}$, and as a result the transformed OLS estimator obtained in the absence of the trend term is not consistent.

A simple way to fix this problem is to use shifted cosine transforms. Let

$$\phi_{Ti}^{c}(r) = \phi_{i}^{c}(r - \frac{1}{2T}) = \sqrt{2}\cos\left(2\pi i(r - \frac{1}{2T})\right) \text{ for } i = 1, ..., K \tag{57}$$

be the finite sample shifted version of $\{\phi_i^c(r)\}_{i=1}^{K}$[1]. We define

$$\breve{\mathbb{W}}_i^v = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_t \phi_{Ti}^c\left(\frac{t}{T}\right) \text{ for } v = y,\ x,\ \Delta x \text{ and} \tag{58}$$

$$\breve{\mathbb{W}}_i^{tr} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} t\phi_{Ti}^c\left(\frac{t}{T}\right),\ \breve{\mathbb{W}}_i^0 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{0t}\phi_{Ti}^c\left(\frac{t}{T}\right). \tag{59}$$

It follows from Lemma 8 in Bierens (1997) that $\breve{\mathbb{W}}_i^{tr} = 0$ for any $i = 1, ..., K$. Also, it is easy to show that $T^{-1/2}\sum_{t=1}^{T}\phi_{Ti}^c(t/T) = 0$ for all $i = 1, ..., K$. So utilizing $\{\phi_{Ti}^c(r)\}_{i=1}^{K}$ as the basis functions filters out both the intercept and linear trend in the original equation (33)[2]. As a result, we have

$$\breve{\mathbb{W}}_i^y = \breve{\mathbb{W}}_i^{x\prime}\beta_0 + \breve{\mathbb{W}}_i^0 \text{ for } i = 1, ..., K. \tag{60}$$

On the basis of this equation, the transformed OLS estimator of β_0 is given by

$$\hat{\beta}_{TOLS} = \left(\breve{\mathbb{W}}^{x\prime}\breve{\mathbb{W}}^{x}\right)^{-1}\breve{\mathbb{W}}^{x\prime}\breve{\mathbb{W}}^{y}.$$

In view of $\phi_{Ti}^c(r) = \phi_i^c(r) + O(1/T)$ uniformly for $r \in [0, 1]$, we have

$$\begin{aligned}
\breve{\mathbb{W}}_i^x &\Rightarrow \Omega_{xx}^{1/2}\eta_i \overset{d}{=} \Omega_{xx}^{1/2}\int_0^1 \phi_i^c(r)\, W_x(r)dr, \\
\breve{\mathbb{W}}_j^{\Delta x} &\Rightarrow \Omega_{xx}^{1/2}\xi_j \overset{d}{=} \Omega_{xx}^{1/2}\int_0^1 \phi_j^c(r)\, dW_x(r),
\end{aligned} \tag{61}$$

for $i, j = 1, ..., K$, and $\breve{\mathbb{W}}_i^x$ and $\breve{\mathbb{W}}_j^{\Delta x}$ are asymptotically independent. Using these and the same proof for Theorem 7, we can prove the theorem below.

Theorem 8 *Consider the model in (33). Let Assumptions 1 and 2 hold. Suppose that the shifted cosine transforms are used. Then under the fixed-K asymptotics,*

$$T(\hat{\beta}_{TOLS} - \beta_0) \Rightarrow MN\left(0, \sigma_0^2\Omega_{xx}^{-1/2}(\eta'\eta)^{-1}\Omega_{xx}^{-1/2}\right).$$

It follows from the theorem and the arguments similar to the proof of Theorem 3 that

$$F_c(\hat{\beta}_{TOLS}) \Rightarrow \frac{K}{K-d}\cdot F_{p,K-d} \text{ and } t_c(\hat{\beta}_{TOLS}) \Rightarrow \sqrt{\frac{K}{K-d}}\cdot t(K-d), \tag{62}$$

where $F_c(\hat{\beta}_{TOLS})$ and $t_c(\hat{\beta}_{TOLS})$ are defined in the same way as in (53) and (54).

[1] The cosine weight functions $\phi_{Ti}^c(t/T)$ are known as Chebishev time polynomials of even orders. See Hamming (1973) for details. Bierens (1997) shows that the cosine basis functions enjoy a certain optimality property for hypothesis testing.

[2] Sun (2011) also uses the cosine basis functions in OS LRV estimation in order to achieve invariance with respect to the intercept and linear trend.

For a cointegration model without a trend, it is not hard to show that Theorem 7 and the asymptotic results for the test statistics thereafter remain the same if we use the shifted cosine transforms in place of the original cosine transforms. That is, for a cointegration model without a trend, it does not matter asymptotically whether the shifted cosine transforms or the original cosine transforms are employed. However, the shifted cosine transforms lead to the TOLS estimator that is invariant to the presence of a linear trend. This is a nice property that is not enjoyed by the original cosine transforms. For this reason, the shifted cosine transforms are preferred over the original cosine transforms.

4.3 Augment or not: asymptotic efficiency comparison

Suppose that we use the shifted cosine transforms. Regardless of whether there is a linear time trend, we have two different estimators of β_0, both of which are asymptotically mixed normal. The first one is the TAOLS estimator and the second one is the TOLS estimator. The difference is whether the underlying regression is augmented or not. In this subsection, we address the relative efficiency of the two estimators.

For the model without a time trend, it follows from (25) and Theorem 7 that the asymptotic variances of $\hat{\beta}_{TAOLS}$ and $\hat{\beta}_{TOLS}$ conditioning on (η, ξ) are

$$V_{TAOLS} = \sigma_{0\cdot x}^2 \Omega_{xx}^{-1/2} (\eta' M_\xi \eta)^{-1} \Omega_{xx}^{-1/2}, \tag{63}$$

$$V_{TOLS} = \sigma_0^2 \Omega_{xx}^{-1/2} (\eta' \eta)^{-1} \Omega_{xx}^{-1/2}, \tag{64}$$

where we call that $\eta = (\eta_1, ..., \eta_K)'$, $\xi = (\xi_1, ..., \xi_K)'$, $\eta_i = \int \phi_i^c(r) W_x(r) dr$ and $\xi_i = \int \phi_i^c(r) dW_x(r)$.

For the model with a linear time trend, we know that $a = 0$ in Theorem 5. So no transformed time trend can be included in the transformed and augmented regression. In this case, we can follow the same proof of Theorem 5 and show that the asymptotic variance of $\hat{\beta}_{TAOLS}$ is $\sigma_{0\cdot x}^2 \Omega_{xx}^{-1/2} (\eta' M_\xi \eta)^{-1} \Omega_{xx}^{-1/2}$. On the other hand, the asymptotic variance of $\hat{\beta}_{TOLS}$ is $\sigma_0^2 \Omega_{xx}^{-1/2} (\eta' \eta)^{-1} \Omega_{xx}^{-1/2}$ as indicated by Theorem 8. That is, the asymptotic variance formulae in (63) and (64) hold regardless of whether a linear trend is included in the model or not.

For any conforming vector $c \in \mathbb{R}^d$, we have

$$\begin{aligned} & c'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c \\ &= \frac{c'\Omega_{xx}^{1/2}}{\sigma_{0\cdot x}} (\eta' M_\xi \eta - \frac{\sigma_{0\cdot x}^2}{\sigma_0^2} \eta'\eta) \frac{\Omega_{xx}^{1/2} c}{\sigma_{0\cdot x}} \\ &= \frac{c'\Omega_{xx}^{1/2}}{\sigma_{0\cdot x}} \left[\eta' (I_K - P_\xi) \eta - \left(\frac{\sigma_0^2 - \sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}}{\sigma_0^2} \right) \eta'\eta \right] \frac{\Omega_{xx}^{1/2} c}{\sigma_{0\cdot x}} \\ &= \frac{c'\Omega_{xx}^{1/2}\eta'}{\sigma_{0\cdot x}} \left[I_K \cdot \left(\frac{\sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}}{\sigma_0^2} \right) - P_\xi \right] \frac{\eta \Omega_{xx}^{1/2} c}{\sigma_{0\cdot x}} \\ &= \tilde{c}' \left[I_K \cdot \varrho^2 - P_\xi \right] \tilde{c} \end{aligned} \tag{65}$$

where $\tilde{c} = \eta \Omega_{xx}^{1/2} c / \sigma_{0\cdot x}$, $P_\xi = \xi(\xi'\xi)^{-1}\xi'$, and

$$\varrho^2 = \frac{\sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}}{\sigma_0^2} = \arg\max_{\ell} \left(\frac{\ell'\sigma_{x0}}{\sqrt{\ell'\Omega_{xx}\ell}\sigma_0} \right)^2 \in [0, 1]. \tag{66}$$

By definition, ϱ^2 is the squared long run canonical correlation coefficient between u_{0t} and u_{xt}. If $\varrho^2 = 0$, then $c'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c = -\tilde{c}' P_\xi \tilde{c} \le 0$ almost surely. In this case, the asymptotic variance of $\hat{\beta}_{TAOLS}$ is always larger than the asymptotic variance of $\hat{\beta}_{TOLS}$. Intuitively, when the long run canonical correlation between u_{0t} and u_{xt} is zero, including the additional regressor $\mathbb{W}^{\Delta x}$ will not help reduce the size of the error term in the transformed regression. However, the presence of $\mathbb{W}^{\Delta x}$ reduces the strength of the signal in $\mathbb{W}^{x}$ even though they are asymptotically independent. That is why $\hat{\beta}_{TAOLS}$ is asymptotically less efficient. On the other hand, when $\varrho^2 = 1$, which holds if the long run variation of u_{0t} can be perfectly predicted by u_{xt}, we have $c'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c = \tilde{c}'\left(I_K - P_{\xi}\right)\tilde{c} \geq 0$ almost surely. In this case, the benefit of including the additional regressor $\mathbb{W}^{\Delta x}$ outweighs the cost, and it is worthwhile to include $\mathbb{W}^{\Delta x}$ to get the asymptotically more efficient estimator $\hat{\beta}_{TAOLS}$.

There are many scenarios between these two extreme cases. Whether the asymptotic distribution of $\hat{\beta}_{TAOLS}$ has a larger variance than that of $\hat{\beta}_{TOLS}$ depends on the value of ϱ^2.

Proposition 9 *If $\varrho^2 \geq d/K$, then $\hat{\beta}_{TAOLS}$ has a smaller asymptotic variance than $\hat{\beta}_{TOLS}$, i.e.,* $\text{asymvar}(\hat{\beta}_{TAOLS}) - \text{asymvar}(\hat{\beta}_{TAOLS})$ *is negative semidefinite. Otherwise, $\hat{\beta}_{TAOLS}$ has a larger asymptotic variance than $\hat{\beta}_{TOLS}$.*

4.4 AMSE Rule

For the cosine basis function $\{\sqrt{2}\cos 2i\pi r\}_{i=1}^{K}$ we can follow Phillips (2005) and Sun (2011) and show that the AMSE-optimal K^* is given by

$$\begin{aligned} K_{MSE}^{c*} &= \left\lceil \left(\frac{1}{16} \frac{tr(I_{m^2} + \mathbb{K}_{mm})(\Omega \otimes \Omega)}{4vec\left(B\right)' vec\left(B\right)} \right)^{1/5} T^{4/5} \right\rceil \\ &\simeq \left(\frac{1}{16}\right)^{1/5} K_{MSE}^{*} = K_{MSE}^{*}(0.57) \end{aligned} \tag{67}$$

where K_{MSE}^{*} is the AMSE-optimal K given in (46) for the basis functions given in (43). Following the same argument for (47), we recommend making an adjustment in finite samples and set K equal to $\max\left(K_{MSE}^{c*}, d+5\right)$.

Given the smaller choice of K, the use of cosine basis functions rather than the complete cosine and sine basis functions may lead to a less efficient estimator of β_0. However, the cosine basis functions enjoy two advantages that the complete basis functions do not. First, it automatically filters out the time trend regressor so that we do not have to worry about the estimation error in trend extraction. Second, the use of cosine basis function renders it unnecessary in some scenarios to include the first difference regressor in the regression and thus saves some degrees of freedom. These two advantages may offset the efficiency loss from having to select a smaller K.

5 Simulation

We compare the finite-sample performance of our method with several existing methods in the literature. For cointegration models without a time trend, we follow Phillips (2014) and consider:

$$\begin{array}{l} y_t = \alpha_0 + x_t'\beta_0 + u_{0t} \\ x_t = x_{t-1} + u_{xt} \end{array}, u_t = \begin{pmatrix} u_{0t} \\ u_{xt} \end{pmatrix} = \Theta u_{t-1} + \epsilon_t \tag{68}$$

where $x_0 = 0$,

$$\epsilon_t = \begin{pmatrix} \epsilon_{0t} \\ \epsilon_{xt} \end{pmatrix} \sim \text{i.i.d } N\left(0, \Sigma\right), \ \Theta = \rho \cdot I_{d+1}, \ \Sigma = J_{d+1,d+1} \cdot \varphi + I_{d+1} \cdot (1 - \varphi),$$

and $J_{p,q}$ is the $p \times q$ matrix of ones. The dimension d of x_t is set to 2, and the true regression coefficients are set to be $\alpha_0 = 3$ and $\beta_0 = (1,1)'$. The parameter ρ controls the persistence of individual components of $u_t = (u_{0t}, u_{xt}')' \in \mathbb{R}^{d+1}$. The parameter φ characterizes the degree of endogeneity, as it is equal to the pairwise correlation coefficient between the elements of u_t in the above model. We set the values of ρ and φ as follows:

$$\rho \in \{0.05, 0.20, 0.35, 0.50, 0.75, 0.90\} \text{ and } \varphi \in \{0, 0.75\}.$$

We also consider a cointegration model that includes a linear time trend:

$$\begin{aligned} y_t &= \alpha_0 + \mu_0 t + x_t'\beta_0 + u_{0t} \\ x_t &= x_{t-1} + u_{xt}, \end{aligned}$$

where μ_0 is set to 0.05 without loss of generality. Other configurations are exactly the same as the model without a linear trend.

We are interested in testing $H_0 : \beta_0 = (1,1)'$ vs. $H_1 : \beta_0 \neq (1,1)'$. We consider the Wald-type tests based on four different estimators: the FMOLS estimator of Phillips and Hansen (1990), the TIV estimator of Phillips (2014), the IMOLS estimator of Vogelsang and Wagner (2014), and the TAOLS estimator proposed in this paper. The first two tests are chi-squared tests that employ the increasing-smoothing asymptotic approximation and use chi-squared critical values. The last two tests are fixed-smoothing tests. The IMOLS test employs the fixed-b asymptotic critical values, which are available from the supplementary appendix to Vogelsang and Wagner (2014). The TAOLS test employs the fixed-K asymptotic approximation and scaled standard F critical values.

For the FMOLS and IMOLS methods, we consider the Bartlett, Parzen, and Quadratic Spectral (QS) kernels. For the TIV method, we choose the basis functions $\{\sqrt{2}\sin((j-1/2)\pi r)\}$, as suggested by Phillips (2014). Note that for the TIV method, a constant vector should be included in the instrument matrix Φ defined in (22). When the model includes a linear time trend, the linear trend should also be included in Φ. For the TAOLS, we consider the Fourier basis functions and cosine functions given in (43) and (44), respectively. Simulation results for models with a linear trend and for shifted cosine basis functions are reported in the original working paper Hwang and Sun (2016).

For fixed values of K, we set $K = 8$ and 16. The comparable values of b for the kernel methods that deliver the same asymptotic variance are

$$b = (c_a K)^{-1} \text{ for } c_a = 2/3, 0.539285, \text{ and } 1,$$

for the Bartlett, Parzen, and QS kernels, respectively. In particular, for the QS kernel, which we will focus on, the corresponding b values are $b = 0.13$ and 0.06.

For data-driven values of K, we employ the formula in either (47) or (48), depending on whether a linear trend is included in the model. For the data-driven values of b, we employ the formulae in Andrews (1991), which are obtained by minimizing the asymptotic (truncated) mean squared error of the kernel LRV estimator. The asymptotic mean squared error criterion is not necessarily the most suitable one for the IMOLS based inference. Ideally, we should derive a formula for b that optimally balances the size distortion under the null and size-corrected power under the alternative. Like the AMSE-based rule given in (47) or (48), the AMSE-based rule of choice for b should be regarded as only a rule of thumb. See Vogelsang and Wagner (2014) for more discussion on the subtlety of choosing b. When data-driven values are used for both K and

b, the unknown parameters B and Ω in the data-driven formulae are estimated by the plug-in method using VAR(1) as the approximating model for $\{\hat{u}_t = (\hat{u}_{0t}^{OLS}, \Delta x_t')'\}$, where $\hat{u}_{0t}^{OLS}$ is the OLS residual based on the OLS estimators of slope coefficients.

We report the simulation evidence only for the model without a linear trend, as the qualitative observations that follow remain valid for the model with a linear trend. Figures 1–3 report the empirical size of five different tests; the labels on the figures should be self-explanatory. The number of simulation replications is 10,000, and the nominal size of all tests is 5%. To avoid overloading the figures, we report only the case with the QS kernel for the kernel-based methods. While Figures 1 and 2 report the case with fixed smoothing parameters, Figure 3 reports the case with data-driven smoothing parameters. Several patterns emerge from these figures.

First, it is clear that, for all values for ρ and φ and sample sizes $T \in \{100, 200\}$, our proposed F tests, i.e., "TA-Fourier-F" and "TA-Cosine-F," which are based on the TAOLS estimator, outperform the chi-squared tests, i.e., "FM-QS-Chi2" and "TIV-Sine-Chi2", by a large margin. Simulation results not reported here show that using F critical values can also dramatically reduce the size distortion of the "TIV-Sine-Chi2" test. Our findings are consistent with the literature on heteroskedasticity-autocorrelation robust inference such as Sun (2013, 2014a), Sun, Phillips, and Jin (2008), and Kiefer and Vogelsang (2005), which provide theoretical justifications and simulation evidence on the accuracy of the fixed-smoothing approximations.

Second, among the two groups of fixed-smoothing tests, our proposed F tests, "TA-Fourier-F" and "TA-Cosine-F," outperform the nonstandard fixed-b test "IM-QS-b." This is true for both the fixed smoothing parameters and data-driven smoothing parameters. In the cases with fixed smoothing parameters, the fixed-smoothing tests are fairly accurate when ρ is small but become somewhat sized distorted when ρ is very large. The exception is that, when $b = 0.13$, the "IM-QS-b" test under-rejects when ρ is small, suggesting that the fixed-b critical value appears to be too large when ρ is small.

Third, the data-driven choices of smoothing parameters help improve the size accuracy of the fixed-smoothing tests. In particular, our proposed F tests are very accurate when K is data-driven. While it is convenient to set the smoothing parameter to a given value, it pays to use a data-driven rule, even though the rule is designed for a different problem. The data-driven rule is more compatible with the increasing-smoothing asymptotics, but the fixed-smoothing critical values are adaptive in that they approach the increasing-smoothing critical value when the amount of smoothing is large. In other words, coupling a data-driven smoothing parameter with fixed-smoothing critical values can be theoretically justified using the increasing-smoothing asymptotics under which $K \to \infty$ as $T \to \infty$.

Fourth, it is not surprising that endogeneity of a higher degree poses more challenges for size accuracy. It is also well expected that when the smoothing parameters are data-driven, a larger sample size helps reduce the size distortion.

Finally, for our proposed F tests, the finite-sample performances are virtually the same across the two sets of trigonometric basis functions we consider. So, in terms of size accuracy, it makes almost no difference which set of trigonometric basis functions is used.

Next, we investigate the finite-sample power of each procedure. The power is size-adjusted so that the comparison is meaningful. The DGP's are the same except that the parameters of interest are from the local alternative hypothesis $\beta = \beta_0 + \theta/T$ where $\theta/\|\theta\|$ is uniform on a sphere. The choice rules for K and b are also the same as before. Each power curve is drawn against $\|\theta\|$, which measures the magnitude of the local departure. Figures 4–8 present the size-adjusted power curve of each procedure for $\rho = 0.05, 0.35, 0.50, 0.75, 0.90$, $\varphi = 0, 0.75$, and

$T = 100, 200$ when the smoothing parameters are data-driven. We omit the TIV test from the figures, as its power curve is close to that of the TA-Fourier test. The closeness of the power curves is not surprising, because the TIV test and the TA-Fourier test are based on essentially the same test statistic. The only difference lies in the critical values used. So upon size adjustment the two tests are essentially the same test.

The simulation results are briefly summarized as follows.

First, the FM-QS test yields the highest power in almost all cases. This is not surprising, as the FMOLS estimator effectively uses both low-frequency and high-frequency components to estimate the cointegrating vector with modification in only the second stage. However, the FMOLS estimator can be fragile if there are high-frequency contaminations. In addition, the FM-QS test has very large size distortion. As an example, for $\varphi = 0.75$ and $T = 200$, the empirical size of the FM-QS test is 25% when $\rho = 0.75$. It increases to 45% when $\rho = 0.90$. These numbers can be read from Figure 3 and are also available from tables not reported here.

Second, the power of the "TA-Fourier" test is either close to or higher than the power of the IM-QS test. When the serial dependence is weak, e.g. $\rho = 0.05$, the "TA-Fourier" test is as powerful as the most powerful FM-QS test. Given its accurate size, superior power, and convenience to use, we recommend the "TA-Fourier-F" test for practical use.

Third, among our proposed F tests, the power curve based on the Fourier bases is close to that based on the cosine bases. While the Fourier bases perform slightly better than the cosine bases when ρ is relatively small, their power curves are virtually indistinguishable when ρ becomes large. In view of the similar size and power properties, we can conclude that it does not matter much whether Fourier bases or cosine bases are used.

To sum up, when the smoothing parameter K is data-driven, the TAOLS based F tests have fairly accurate size. They are much more accurate than the FM-QS and TIV tests that use the chi-squared approximation. They are also more accurate than the IM-QS test, which also uses a fixed-smoothing approximation. While the TAOLS based F tests are not as powerful as the FM-QS test for some simulation configurations, they have competitive and often superior power relative to the IM-QS test.

6 Conclusion

This paper provides a simple, robust, and more accurate approach to parameter estimation and inference in a triangular cointegrated system. Cointegration is fundamentally a long run relationship. Our approach echoes this key observation by focusing only on data transformations that capture the long run variation and covariation of the underlying time series. From a practical point of view, our approach enjoys two major advantages. First, the more accurate approximations we derived under the so-called fixed-K asymptotics are the standard F and t distributions. Second, test statistics can be obtained from the usual regression output. So, our asymptotic F and t tests are just as easy to implement as the F and t tests in a classical linear normal regression. A simulation study shows that our tests are much more accurate than the chi-squared tests. For practical use, we recommend using the Fourier basis functions and employing the modified data-driven rule to select the number of basis functions.

A key open question is how to select the number of basis functions optimally. While we have suggested a data-driven approach, it does not directly target the problem under consideration. It would be interesting to investigate selection of the number of basis functions to minimize the approximate mean squared error of the point estimator of the cointegrating vector. If we are interested in interval estimation or hypothesis testing, then the number of basis functions should be oriented toward optimizing the underlying objects such as the coverage probability error, the interval length, and the type I and type II errors. There may also be room to select optimal basis functions. We hope to address some of these questions in future research.

7 Appendix of Proofs

Proof of Theorem 1. By the definition of $\hat{\gamma}_{TAOLS}$ and Υ_T, we have

$$\Upsilon_T\left(\hat{\gamma}_{TAOLS}-\gamma_0\right)=(\Upsilon_T^{-1}\tilde{\mathbb{W}}'\tilde{\mathbb{W}}\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}\tilde{\mathbb{W}}'\mathbb{W}^{0\cdot x}. \tag{69}$$

Note that $\tilde{\mathbb{W}}\Upsilon_T^{-1}=\left(\mathbb{W}^x/T,\,\mathbb{W}^{\Delta x}\right)$ where

$$\mathbb{W}^x/T=(\mathbb{W}_1^x/T,...,\mathbb{W}_K^x/T)'=\left(\frac{1}{T^{3/2}}\sum_{t=1}^T\phi_1(\frac{t}{T})x_t,...,\frac{1}{T^{3/2}}\sum_{t=1}^T\phi_K(\frac{t}{T})x_t\right)'$$

and

$$\mathbb{W}^{\Delta x}=\left(\mathbb{W}_1^{\Delta x},...,\mathbb{W}_K^{\Delta x}\right)'=\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_1(\frac{t}{T})u_{xt},...,\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_K(\frac{t}{T})u_{xt}\right)'.$$

By Assumption 1 and the continuous mapping theorem,

$$\frac{1}{T^{3/2}}\sum_{t=1}^T\phi_i(\frac{t}{T})x_t\Rightarrow\Omega_{xx}^{1/2}\left(\int_0^1\phi_i(r)W_x(r)dr\right)=\Omega_{xx}^{1/2}\eta_i, \tag{70}$$

$$\text{and }\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi_j(\frac{t}{T})u_{xt}\Rightarrow\Omega_{xx}^{1/2}\left(\int_0^1\phi_j(r)dW_x(r)\right):=\Omega_{xx}^{1/2}\xi_j, \tag{71}$$

hold jointly over $i,j=1,...,K$. So

$$\mathbb{W}^x/T\Rightarrow(\Omega_{xx}^{1/2}\eta')'\text{ and }\mathbb{W}^{\Delta x}\Rightarrow(\Omega_{xx}^{1/2}\xi')',$$

and

$$\tilde{\mathbb{W}}\Upsilon_T^{-1}\Rightarrow\left((\Omega_{xx}^{1/2}\eta')',(\Omega_{xx}^{1/2}\xi')'\right)=\left(\eta\Omega_{xx}^{1/2},\xi\Omega_{xx}^{1/2}\right). \tag{72}$$

Similarly, we have

$$\begin{aligned}\mathbb{W}^{0\cdot x}&=\left(\frac{1}{T^{1/2}}\sum_{t=1}^T\phi_1(\frac{t}{T})u_{0\cdot xt},...,\frac{1}{T^{1/2}}\sum_{t=1}^T\phi_K(\frac{t}{T})u_{0\cdot xt}\right)'\\&\Rightarrow(\sigma_{0\cdot x}\nu_1,\sigma_{0\cdot x}\nu_2,...,\sigma_{0\cdot x}\nu_K)'=\sigma_{0\cdot x}\nu\end{aligned} \tag{73}$$

where $\nu=[\nu_1,...,\nu_K]'\sim N(0,I_K)$. The above convergence holds jointly with (72), i.e.,

$$\left(\tilde{\mathbb{W}}\Upsilon_T^{-1},\mathbb{W}^{0\cdot x}\right)\Rightarrow\left(\tilde{\zeta},\tilde{\nu}\right)\text{ where }\tilde{\zeta}=\left(\eta\Omega_{xx}^{1/2},\xi\Omega_{xx}^{1/2}\right),\ \tilde{\nu}:=\sigma_{0\cdot x}\nu\text{ and }\tilde{\zeta}\perp\tilde{\nu}. \tag{74}$$

Using this result, we have

$$\Upsilon_T\left(\hat{\gamma}_{TAOLS}-\gamma_0\right)=(\Upsilon_T^{-1}\tilde{\mathbb{W}}'\tilde{\mathbb{W}}\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}\tilde{\mathbb{W}}'\mathbb{W}^{0\cdot x}$$
$$\Rightarrow(\tilde{\zeta}'\tilde{\zeta})^{-1}\tilde{\zeta}'\tilde{\nu}\overset{d}{=}MN\left[0,\sigma_{0\cdot x}^2(\tilde{\zeta}'\tilde{\zeta})^{-1}\right].$$

The weak limit can be written more explicitly as

$$\begin{aligned}(\tilde{\zeta}'\tilde{\zeta})^{-1}\tilde{\zeta}'\tilde{\nu}&=\sigma_{0\cdot x}\begin{pmatrix}\Omega_{xx}^{1/2}\eta'\eta\Omega_{xx}^{1/2} & \Omega_{xx}^{1/2}\eta'\xi\Omega_{xx}^{1/2}\\ \Omega_{xx}^{1/2}\xi'\eta\Omega_{xx}^{1/2} & \Omega_{xx}^{1/2}\xi'\xi\Omega_{xx}^{1/2}\end{pmatrix}^{-1}\begin{pmatrix}\Omega_{xx}^{1/2}\eta'\\ \Omega_{xx}^{1/2}\xi'\end{pmatrix}\nu\\
&=\sigma_{0\cdot x}\begin{pmatrix}\Omega_{xx}^{-1/2} & 0\\ 0 & \Omega_{xx}^{-1/2}\end{pmatrix}\begin{pmatrix}\eta'\eta & \eta'\xi\\ \xi'\eta & \xi'\xi\end{pmatrix}^{-1}\begin{pmatrix}\eta'\\ \xi'\end{pmatrix}\nu\\
&=\begin{pmatrix}\sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\eta'M_\xi\eta\right)^{-1}\eta'M_\xi\nu\\ \sigma_{0\cdot x}\Omega_{xx}^{-1/2}\left(\xi'M_\eta\xi\right)^{-1}\xi'M_\eta\nu\end{pmatrix}.\end{aligned}\tag{75}$$

So the representations in (25) and (26) hold. ■

Proof of of Proposition 2. Part (i). By Lemma A in Section 6 of Phillips (2005b), we have

$$\Phi'\Phi=I_K+O\left(\frac{1}{T}\right)\text{ and }(\Phi'\Phi)^{-1}=I_K+O\left(\frac{1}{T}\right)$$

for any fixed K. Then it is straightforward to show that

$$P_\Phi=\Phi\left(\Phi'\Phi\right)^{-1}\Phi'=\Phi\left(I_K+O(T^{-1})\right)\Phi'=\Phi\Phi'+O(T^{-1}).$$

Now

$$T(\hat{\beta}_{TIV}-\beta_0)=(X'S_\Phi X)^{-1}(X'S_\Phi U^{0\cdot x})=\left(\frac{X'}{T^{3/2}}S_\Phi\frac{X}{T^{3/2}}\right)^{-1}\left(\frac{X'}{T^{3/2}}S_\Phi\frac{U^{0\cdot x}}{T^{1/2}}\right),\tag{76}$$

where $U^{0\cdot x}=(u_{0\cdot x,1},\ldots,u_{0\cdot x,T})'$ and $S_\Phi=P_\Phi-P_\Phi\Delta X(\Delta X'P_\Phi\Delta X)^{-1}\Delta X'P_\Phi$. Note that

$$\begin{aligned}S_\Phi&=\Phi\left\{\left(\Phi\Phi'\right)^{-1}-\left(\Phi\Phi'\right)^{-1}\frac{\Phi'\Delta X}{\sqrt{T}}\left[\left(\frac{\Delta X'\Phi}{\sqrt{T}}\right)\left(\Phi\Phi'\right)^{-1}\left(\frac{\Phi'\Delta X}{\sqrt{T}}\right)\right]^{-1}\frac{\Delta X'\Phi}{\sqrt{T}}\left(\Phi\Phi'\right)^{-1}\right\}\Phi'\\
&=\Phi\Phi'+O(T^{-1})\\
&-\Phi\left[I_K+O\left(T^{-1}\right)\right]\mathbb{W}^{\Delta x}\left\{\mathbb{W}^{\Delta x\prime}\left[I_K+O\left(T^{-1}\right)\right]\mathbb{W}^{\Delta x}\right\}^{-1}\mathbb{W}^{\Delta x\prime}\left[I_K+O\left(T^{-1}\right)\right]\Phi'\\
&=\Phi\left(I_K-P_{\Delta x}\right)\Phi'+o_p(1),\end{aligned}$$

hence we have

$$\begin{aligned}\frac{X'}{T^{3/2}}S_\Phi\frac{X}{T^{3/2}}&=\frac{X'\Phi}{T^{3/2}}\left(I_K-P_{\Delta x}+o_p(1)\right)\frac{\Phi'X}{T^{3/2}}\\
&=\frac{\mathbb{W}^{x\prime}}{T}\left(I_K-P_{\Delta x}\right)\frac{\mathbb{W}^x}{T}+o_p(1)\end{aligned}$$

and

$$\begin{aligned}\frac{X'}{T^{3/2}}S_\Phi\frac{U^{0\cdot x}}{T^{1/2}}&=\left(\frac{X'\Phi}{T^{3/2}}\right)'\left(I_K-P_{\Delta x}+o_p(1)\right)\frac{\Phi'U^{0\cdot x}}{\sqrt{T}}\\
&=\frac{\mathbb{W}^{x\prime}}{T}\left(I_K-P_{\Delta x}\right)\mathbb{W}^{0\cdot x}+o_p(1).\end{aligned}\tag{77}$$

Combining these two representations and using Theorem 1, we have

$$
\begin{aligned}
T(\hat{\beta}_{TIV}-\beta_0) &= \left(\frac{\mathbb{W}^{x\prime}}{T}\left(I_K-P_{\Delta x}\right)\frac{\mathbb{W}^{x}}{T}\right)^{-1}\left(\frac{\mathbb{W}^{x\prime}}{T}\left(I_K-P_{\Delta x}\right)\right)\mathbb{W}^{0\cdot x}+o_p(1)\\
&= T(\hat{\beta}_{TAOLS}-\beta_0)+o_p(1).
\end{aligned}
$$

Part (ii): For any conformable vector c and $\tau\in\mathbb{R}$, we have

$$
\lim_{K\to\infty}P\left(c'V_K<\tau\right)=\lim_{K\to\infty}EG\left(\frac{\tau}{\sigma_{0\cdot x}\sqrt{c'\left[\Omega_{xx}^{-1/2}(\eta'M_\xi\eta)^{-1}\Omega_{xx}^{-1/2}\right]c}}\right),
$$

where $G(\cdot)$ is the cdf of the standard normal distribution. Note that $\int_0^1 W_x(r)W_x(r)'dr$ is positive definite with probability one. By the definition of weak convergence and the continuous mapping theorem, it suffices to show that

$$
\eta'M_\xi\eta\Rightarrow\int_0^1\tilde{W}_x(r)\tilde{W}_x(r)'dr\text{ as }K\to\infty.
$$

We first show that $\eta'\eta\Rightarrow\int_0^1\tilde{W}_x(r)\tilde{W}_x(r)'dr$. Given that $\{\phi_i(\cdot)\}_{i=1}^{\infty}$ is a complete orthonormal system in $L_0^2[0,1]$, we know that $\sum_{i=1}^K\phi_i(r)\phi_i(s)$ converges to the Dirac delta function $\delta(r-s)$ in that

$$
\left\|\int_0^1\left(\sum_{i=1}^K\phi_i(r)\phi_i(s)\right)f(r)dr-f(s)\right\|_{L^2}\to 0 \tag{78}
$$

for any $f\in L_0^2[0,1]\cap C[0,1]$ as $K\to\infty$ where $\|\cdot\|_{L^2}$ is the L^2 norm and $C[0,1]$ is the space of continuous functions on $[0,1]$. But

$$
\begin{aligned}
\eta'\eta=\sum_{i=1}^K\eta_i\eta_i' &= \sum_{i=1}^K\left(\int_0^1\phi_i(r)W_x(r)dr\right)\left(\int_0^1\phi_i(s)W_x(s)ds\right)'\\
&=\int_0^1\int_0^1\left(\sum_{i=1}^K\phi_i(r)\phi_i(s)\right)\tilde{W}_x(r)\tilde{W}_x'(s)drds
\end{aligned}
$$

where $\tilde{W}_x(s)\in L_0^2[0,1]\cap C[0,1]$ almost surely, and so

$$
\eta'\eta\to\int_0^1\tilde{W}_x(r)\tilde{W}_x'(r)dr \tag{79}
$$

almost surely. This implies that $\eta'\eta\Rightarrow\int_0^1\tilde{W}_x(r)\tilde{W}_x(r)'dr$.

Next, we prove $\eta'P_\xi\eta=o_p(1)$. We have

$$
\begin{aligned}
\eta'P_\xi\eta=\eta'\xi\left(\xi'\xi\right)^{-1}\xi'\eta &= \frac{1}{K}\cdot\eta'\xi\left(\frac{\xi'\xi}{K}\right)^{-1}\xi'\eta\\
&=\frac{1}{K}\cdot\left(\sum_{i=1}^K\eta_i\xi_i'\right)\left(\frac{1}{K}\sum_{i=1}^K\xi_i\xi_i'\right)^{-1}\left(\sum_{i=1}^K\xi_i\eta_i'\right)\\
&=\frac{1}{K}\cdot\left(\sum_{i=1}^K\eta_i\xi_i'\right)\left(\sum_{i=1}^K\xi_i\eta_i'\right)+o_p(1)
\end{aligned} \tag{80}
$$

where the last equality follows from the result that $\xi_i \overset{i.i.d}{\sim} N(0, I_d)$ for $i = 1, \ldots, K$. For the term $\sum_{i=1}^{K} \eta_i \xi_i'$, we note that

$$\begin{aligned}
\eta_i &= \int_0^1 \phi_i(r) W_x(r) dr = \int_0^1 \phi_i(r) \left[\int_0^r dW_x(s) \right] dr = \int_0^1 \left(\int_s^1 \phi_i(r) dr \right) dW_x(s) \\
&= - \int_0^1 \left(\int_0^s \phi_i(r) dr \right) dW_x(s) = - \int_0^1 \Psi_i(s) dW_x(s) \qquad (81)
\end{aligned}$$

where $\Psi_i(s) = \int_0^s \phi_i(r) dr$. So

$$\begin{aligned}
E \sum_{i=1}^{K} \eta_i \xi_i' &= - \sum_{i=1}^{K} E \left[\int_0^1 \Psi_i(s) dW_x(s) \right] \left[\int_0^1 \phi_i(r) dW_x'(r) \right] \\
&= -I_d \sum_{i=1}^{K} \int_0^1 \Psi_i(r) \phi_i(r) dr = -I_d \int_0^1 \sum_{i=1}^{K} \Psi_i(r) \phi_i(r) dr \\
&= -I_d \int_0^1 \left(\int_0^1 \sum_{i=1}^{K} \phi_i(s) \{ s \leq r \} ds \right) \phi_i(r) dr \\
&= -I_d \int_0^1 \left[\int_s^1 \sum_{i=1}^{K} \phi_i(s) \phi_i(r) dr \right] ds \\
&= -I_d \int_0^1 \left[\int_0^{1-s} \sum_{i=1}^{K} \phi_i(s) \phi_i(r+s) dr \right] ds \\
&\to -I_d \int_0^1 \left[\int_0^1 \delta(r) dr \right] ds = O(1),
\end{aligned}$$

and

$$\begin{aligned}
& var \left[vec \left(\sum_{i=1}^{K} \eta_i \xi_i' \right) \right] \\
&= var \left[\sum_{i=1}^{K} \int_0^1 \int_0^1 \Psi_i(r) \phi_i(s) vec \left(dW_x(r) dW_x'(s) \right) \right] \\
&= var \left[\sum_{i=1}^{K} \int_0^1 \int_0^1 \Psi_i(r) \phi_i(s) \left(dW_x(s) \otimes dW_x(r) \right) \right] \\
&= \sum_{i=1}^{K} \sum_{j=1}^{K} E \left[\int_0^1 \int_0^1 \int_0^1 \int_0^1 \Psi_i(r) \phi_i(s) \Psi_j(p) \phi_j(q) \left[dW_x(s) \otimes dW_x(r) \right] \left[dW_x'(q) \otimes dW_x'(p) \right] \right] \\
&= \sum_{i=1}^{K} \sum_{j=1}^{K} \left[\int_0^1 \int_0^1 \Psi_i(r) \phi_i(r) \Psi_j(p) \phi_j(p) vec(I_d) vec(I_d)' dr dp \right] \\
&+ \sum_{i=1}^{K} \sum_{j=1}^{K} \int_0^1 \int_0^1 \Psi_i(r) \phi_i(s) \Psi_j(r) \phi_j(s) dr ds \left[I_d \otimes I_d \right] \left(I_{d^2} + \mathbb{K}_{d,d} \right)
\end{aligned}$$

where $\mathbb{K}_{d,d}$ is the $d^2 \times d^2$ commutation matrix. Now

$$\sum_{i=1}^{K}\sum_{j=1}^{K}\left[\int_0^1\int_0^1 \Psi_i(r)\phi_i(r)dr\int_0^1 \Psi_j(p)\phi_j(p)dp\right] vec(I_d)vec(I_d)' = 0$$

because $\int_0^1 \Psi_i(r)\phi_i(r)dr = \frac{1}{2}\left[\Psi_i(r)\right]^2\Big|_0^1 = 0$. Also

$$\begin{aligned}
&\int_0^1\int_0^1 \Psi_i(r)\phi_i(s)\Psi_j(r)\phi_j(s)drds \\
&= \int_0^1 \Psi_i(r)\Psi_j(r)dr\int_0^1 \phi_i(s)\phi_j(s)ds = 1\{i=j\}\int_0^1 \left[\Psi_i(r)\right]^2 dr \\
&= 1\{i=j\}\int_0^1\left[\int_0^r \phi_i(s)ds\right]^2 dr. \qquad (82)
\end{aligned}$$

As a result, we have

$$\begin{aligned}
&var\left[vec\left(\sum_{i=1}^{K}\eta_i\xi_i'\right)\right] \\
&= \sum_{i=1}^{K}\int_0^1\left[\int_0^r \phi_i(s)ds\right]^2 dr \times \left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) \\
&= \int_0^1\int_0^1\int_0^1\sum_{i=1}^{K}\phi_i(p)\phi_i(q)1\{p\le r\}1\{q\le r\}dpdqdr \times \left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) \\
&\to \int_0^1\int_0^1 1\{p\le r\}dpdr \times \left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) \\
&= \int_0^1 rdr \times \left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) = \frac{1}{2}\left[I_d\otimes I_d\right]\left(I_{d^2}+\mathbb{K}_{d,d}\right) \qquad (83)
\end{aligned}$$

as $K\to\infty$. In view of the mean and variance orders, we have $\sum_{i=1}^{K}\eta_i\xi_i' = O_p(1)$. It then follows that

$$\begin{aligned}
\eta' P_\xi \eta &= \frac{1}{K}\cdot\left(\sum_{k=1}^{K}\eta_i\xi_i'\right)\left(\sum_{k=1}^{K}\xi_i\eta_i'\right) + o_p(1) \\
&= O_p(\frac{1}{K}) + o_p(1) = o_p(1). \qquad (84)
\end{aligned}$$

Combining (79) and (84) yields

$$\eta' M_\xi \eta = \eta'\eta - \eta' P_\xi \eta = \eta'\eta + o_p(1) \Rightarrow \int_0^1 \tilde{W}_x(r)\tilde{W}_x(r)'dr$$

as desired. ■

Proof of of Theorem 3. We prove only the result for the Wald statistic as the proof goes through for the t statistic with obvious modifications. Using (74), we have

$$\begin{aligned}
\hat{\sigma}^2_{0\cdot x} &= \frac{1}{K}\mathbb{W}^{0\cdot x\prime}\left[I_K - \tilde{\mathbb{W}}\left(\tilde{\mathbb{W}}'\tilde{\mathbb{W}}\right)^{-1}\tilde{\mathbb{W}}'\right]\mathbb{W}^{0\cdot x} \\
&= \frac{1}{K}\mathbb{W}^{0\cdot x\prime}\left\{I_K - \tilde{\mathbb{W}}\Upsilon_T^{-1}\left[\left(\tilde{\mathbb{W}}\Upsilon_T^{-1}\right)'\tilde{\mathbb{W}}\Upsilon_T^{-1}\right]^{-1}\left(\tilde{\mathbb{W}}\Upsilon_T^{-1}\right)'\right\}\mathbb{W}^{0\cdot x} \\
&\Rightarrow \sigma^2_{0\cdot x}\frac{1}{K}\nu'\left[I_K - \tilde{\zeta}\left(\tilde{\zeta}'\tilde{\zeta}\right)^{-1}\tilde{\zeta}'\right]\nu.
\end{aligned}$$

But

$$\begin{aligned}
&\tilde{\zeta}\left(\tilde{\zeta}'\tilde{\zeta}\right)^{-1}\tilde{\zeta}' \\
&= \left(\eta\Omega_{xx}^{1/2}, \xi\Omega_{xx}^{1/2}\right)\begin{pmatrix} \Omega_{xx}^{1/2}\eta'\eta\Omega_{xx}^{1/2} & \Omega_{xx}^{1/2}\eta'\xi\Omega_{xx}^{1/2} \\ \Omega_{xx}^{1/2}\xi'\eta\Omega_{xx}^{1/2} & \Omega_{xx}^{1/2}\xi'\xi\Omega_{xx}^{1/2}\end{pmatrix}^{-1}\begin{pmatrix}\Omega_{xx}^{1/2}\eta' \\ \Omega_{xx}^{1/2}\xi'\end{pmatrix} \\
&= \left(\eta\Omega_{xx}^{1/2}, \xi\Omega_{xx}^{1/2}\right)\left[\begin{pmatrix}\Omega_{xx}^{1/2} & 0 \\ 0 & \Omega_{xx}^{1/2}\end{pmatrix}\begin{pmatrix}\eta'\eta & \eta'\xi \\ \xi'\eta & \xi'\xi\end{pmatrix}\begin{pmatrix}\Omega_{xx}^{1/2} & 0 \\ 0 & \Omega_{xx}^{1/2}\end{pmatrix}\right]^{-1}\begin{pmatrix}\Omega_{xx}^{1/2}\eta' \\ \Omega_{xx}^{1/2}\xi'\end{pmatrix} \\
&= \left(\eta\Omega_{xx}^{1/2}, \xi\Omega_{xx}^{1/2}\right)\begin{pmatrix}\Omega_{xx}^{-1/2} & 0 \\ 0 & \Omega_{xx}^{-1/2}\end{pmatrix}\begin{pmatrix}\eta'\eta & \eta'\xi \\ \xi'\eta & \xi'\xi\end{pmatrix}^{-1}\begin{pmatrix}\Omega_{xx}^{-1/2} & 0 \\ 0 & \Omega_{xx}^{-1/2}\end{pmatrix}\begin{pmatrix}\Omega_{xx}^{1/2}\eta' \\ \Omega_{xx}^{1/2}\xi'\end{pmatrix} \\
&= (\eta, \xi)\begin{pmatrix}\eta'\eta & \eta'\xi \\ \xi'\eta & \xi'\xi\end{pmatrix}^{-1}\begin{pmatrix}\eta' \\ \xi'\end{pmatrix} := P_\zeta,
\end{aligned}$$

where $\zeta = (\eta, \xi) \in \mathbb{R}^{K\times 2d}$, so $\hat{\sigma}^2_{0\cdot x} \Rightarrow \sigma^2_{0\cdot x}\frac{1}{K}\nu' M_\zeta \nu$ for $M_\zeta = I_K - P_\zeta$. Combining this with

$$T\left(R\hat{\beta}_{TAOLS} - r\right) \Rightarrow \sigma_{0\cdot x}R\Omega_{xx}^{-1/2}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu$$

and

$$\begin{aligned}
&R\left(\frac{\mathbb{W}^{x\prime}}{T}M_{\Delta x}\frac{\mathbb{W}^{x}}{T}\right)^{-1}R' \\
&= R\left\{\frac{\mathbb{W}^{x\prime}}{T}\left[I_K - \mathbb{W}^{\Delta x}(\mathbb{W}^{\Delta x\prime}\mathbb{W}^{\Delta x})^{-1}\mathbb{W}^{\Delta x\prime}\right]\frac{\mathbb{W}^{x}}{T}\right\}^{-1}R' \\
&\Rightarrow R\left\{\Omega_{xx}^{1/2}\eta'\left[I_K - \xi\Omega_{xx}^{1/2}\left(\Omega_{xx}^{1/2}\xi'\xi\Omega_{xx}^{1/2}\right)^{-1}\Omega_{xx}^{1/2}\xi'\right]\eta\Omega_{xx}^{1/2}\right\}^{-1}R' \\
&= R\left\{\Omega_{xx}^{1/2}\eta'\left(I_K - \xi\left(\xi'\xi\right)^{-1}\xi'\right)\eta\Omega_{xx}^{1/2}\right\}^{-1}R' \\
&= R\Omega_{xx}^{-1/2}\left[\eta' M_\xi \eta\right]^{-1}\Omega_{xx}^{-1/2}R', \qquad (85)
\end{aligned}$$

we have

$$\begin{aligned}
F(\hat{\beta}_{TAOLS}) &\Rightarrow \frac{K}{p}\frac{\left[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu\right]'\left(\tilde{R}\left[\eta' M_\xi \eta\right]^{-1}\tilde{R}'\right)^{-1}\left[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu\right]}{\nu' M_\zeta \nu} \\
&= \frac{K}{p}\frac{Q'\left(\tilde{R}\eta' M_\xi \eta\tilde{R}'\right)^{-1}Q}{\nu' M_\zeta \nu}, \qquad (86)
\end{aligned}$$

where $\tilde{R} = R\Omega_{xx}^{1/2}$,

$$Q'\left(\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'\right)^{-1} Q \overset{d}{=} \chi_p^2 \text{ and } \nu' M_\zeta \nu \overset{d}{=} \chi_{K-2d}^2.$$

Note that conditional on $\zeta = (\eta, \xi)$, $M_\zeta \nu = \left[I_K - \zeta\left(\zeta'\zeta\right)^{-1}\zeta'\right]\nu$ and $\eta' M_\xi \nu$ are independent, as both $M_\zeta \nu$ and $\eta' M_\xi \nu$ are normal and the conditional covariance is

$$cov\left(M_\zeta \nu, \eta' M_\xi \nu\right) = \left[I_K - \zeta\left(\zeta'\zeta\right)^{-1}\zeta'\right] M_\xi \eta = 0.$$

So conditional on ζ, the numerator and the denominator in (86) are independent chi-square variates. This implies that

$$\frac{K}{p}\frac{Q'\left(\tilde{R}\left[\eta' M_\xi \eta\right]^{-1}\tilde{R}'\right)^{-1} Q}{\nu' M_\zeta \nu} = \frac{K}{K-2d}\frac{Q'\left(\tilde{R}\left[\eta' M_\xi \eta\right]^{-1}\tilde{R}'\right)^{-1} Q/p}{\nu' M_\zeta \nu/(K-2d)} \overset{d}{=} \frac{K}{K-2d}F_{p,K-2d}$$

conditional on ζ. But the conditional distribution does not depend on the conditioning variable ζ, so it is also the unconditional distribution. We have therefore proved that

$$F(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{K-2d}F_{p,K-2d}. \tag{87}$$

■

Proof of Theorem 4. The proof is similar to that for Theorem 3. We prove the result for the F statistic only. We still have

$$\hat{\sigma}_{0\cdot x}^2 \Rightarrow \sigma_{0\cdot x}^2 \frac{1}{K}\nu' M_\zeta \nu \text{ for } M_\zeta = I_K - P_\zeta$$

and

$$R\left(\frac{\mathbb{W}^{x\prime}}{T} M_{\Delta x}\frac{\mathbb{W}^x}{T}\right)^{-1} R' \Rightarrow R\Omega_{xx}^{-1/2}\left(\eta' M_\xi \eta\right)^{-1}\Omega_{xx}^{-1/2} R'. \tag{88}$$

But now

$$T\left(R\hat{\beta}_{TAOLS} - r\right) \Rightarrow \sigma_{0\cdot x}\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu + \theta,$$

so

$$\begin{aligned} F(\hat{\beta}_{TAOLS}) &\Rightarrow \frac{K}{p}\frac{\left[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu + \frac{\theta}{\sigma_{0\cdot x}}\right]'\left(\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'\right)^{-1}\left[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu + \frac{\theta}{\sigma_{0\cdot x}}\right]}{\nu' M_\zeta \nu} \\ &= \frac{K}{p}\frac{\left\|Z + \left[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'\right]^{-1/2}\theta/\sigma_{0\cdot x}\right\|^2}{\nu' M_\zeta \nu}, \end{aligned} \tag{89}$$

where

$$Z = \left[\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\tilde{R}'\right]^{-1/2}\tilde{R}\left(\eta' M_\xi \eta\right)^{-1}\eta' M_\xi \nu \sim N(0, I_p)$$

and Z is independent of $\nu' M_\zeta \nu$ conditional on $\zeta = (\eta, \xi)$. Using the same conditioning argument as in the proof of Theorem 3, we have

$$\frac{K}{p} \frac{\left\| Z + \left[\tilde{R} \left(\eta' M_\xi \eta \right)^{-1} \tilde{R}' \right]^{-1/2} \theta / \sigma_{0 \cdot x} \right\|^2}{\nu' M_\zeta \nu}$$
$$\overset{d}{=} \frac{K}{K - 2d} F_{p, K-2d} \left(\frac{\theta' \left[\tilde{R} \left(\eta' M_\xi \eta \right)^{-1} \tilde{R}' \right]^{-1} \theta}{\sigma_{0 \cdot x}^2} \right)$$

conditionally on $\zeta = (\eta, \xi)$. Unconditionally, the limiting distribution is a mixed noncentral F distribution with a random noncentrality parameter $\theta' \left[\tilde{R} \left(\eta' M_\xi \eta \right)^{-1} \tilde{R}' \right]^{-1} \theta / \sigma_{0 \cdot x}^2$, which is equal to $\|\lambda\|^2$. ■

Proof of Theorem 5. We follow the same step as in the proof of Theorem 1. We consider only $F_{tr}(\hat{\beta}_{TAOLS})$. The proof for $t_{tr}(\hat{\beta}_{TAOLS})$ is similar.

Let

$$\Upsilon_{T,tr} = \begin{pmatrix} \Upsilon_T & 0 \\ 0 & T^{3/2} \end{pmatrix}$$

Then

$$\Upsilon_{T,tr} \begin{pmatrix} \hat{\beta}_{TAOLS} - \beta_0 \\ \hat{\delta}_{TAOLS} - \delta_0 \\ \hat{\mu}_{TAOLS} - \mu_0 \end{pmatrix} = (\Upsilon_{T,tr}^{-1} \tilde{\mathbb{W}}_{tr}' \tilde{\mathbb{W}}_{tr} \Upsilon_{T,tr}^{-1})^{-1} \Upsilon_{T,tr} \tilde{\mathbb{W}}_{tr}' \mathbb{W}^y \Rightarrow \left(\tilde{\zeta}_{tr}' \tilde{\zeta}_{tr} \right)^{-1} \tilde{\zeta}_{tr}' \tilde{\nu},$$

where

$$\tilde{\zeta}_{tr} = \left(\eta \Omega_{xx}^{1/2}, \xi \Omega_{xx}^{1/2}, a \right).$$

Some simple calculations show that

$$\left(\tilde{\zeta}_{tr}' \tilde{\zeta}_{tr} \right)^{-1} \tilde{\zeta}_{tr}' \tilde{\nu} = \begin{pmatrix} \sigma_{0 \cdot x} \Omega_{xx}^{-1/2} \left(\eta' M_{\xi,a} \eta \right)^{-1} \eta' M_{\xi,a} \nu \\ \sigma_{0 \cdot x} \Omega_{xx}^{-1/2} \left(\xi' M_{\eta,a} \xi \right)^{-1} \xi' M_{\eta,a} \nu \\ \sigma_{0 \cdot x} \left(a' M_{\eta,\xi} a \right)^{-1} a' M_{\eta,\xi} \nu \end{pmatrix}.$$

So part (i) of the theorem holds. In particular,

$$T(\hat{\beta}_{TAOLS} - \beta_0) \Rightarrow \sigma_{0 \cdot x} \Omega_{xx}^{-1/2} \left(\eta' M_{\xi,a} \eta \right)^{-1} \eta' M_{\xi,a} \nu. \tag{90}$$

Following the same steps in the proof of Theorem 3, we have $(\hat{\sigma}_{0 \cdot x}^{tr})^2 \Rightarrow \sigma_{0 \cdot x}^2 \frac{1}{K} \nu' M_{\zeta,a} \nu$. Combining this with (90), we have

$$F_{tr}(\hat{\beta}_{TAOLS}) \Rightarrow \frac{K}{p} \frac{\left[\tilde{R} \left(\eta' M_{\xi,a} \eta \right)^{-1} \eta' M_{\xi,a} \nu \right]' \left(\tilde{R} \left[\eta' M_{\xi,a} \eta \right]^{-1} \tilde{R}' \right)^{-1} \left[\tilde{R} \left(\eta' M_{\xi,a} \eta \right)^{-1} \eta' M_{\xi,a} \nu \right]}{\nu' M_{\zeta,a} \nu}$$
$$=^d \frac{K}{K - 2d - 1} F_{p, K-2d-1}. \tag{91}$$

■

Proof of Lemma 6. The cosine functions clearly satisfy Assumption 1. Note that

$$\Psi_i^c(r) := \int_0^r \phi_i^c(s)\,ds = \sqrt{2}\int_0^r \cos(2i\pi s)ds$$
$$= \sqrt{2}\,\frac{\sin(2i\pi s)}{2i\pi}\bigg|_0^r = \frac{\sin(2i\pi r)}{\sqrt{2}i\pi},$$

we have, for $i \neq j$,

$$\begin{aligned}\int_0^1 \Psi_i^c(r)\phi_j^c(r)\,dr &= \int_0^1 \frac{\sin(2i\pi r)}{\sqrt{2}i\pi}\cdot\sqrt{2}\cos(2j\pi r)\,dr \\ &= \frac{1}{i\pi}\int_0^1 \sin(2i\pi r)\cos(2j\pi r)dr \\ &= \frac{1}{2i\pi}\left(\int_0^1 \sin(2(i+j)\pi r)\,dr + \int_0^1 \sin(2(i-j)\pi r)\,dr\right) \\ &= \frac{1}{2i\pi}\left(-\frac{\cos(2(i+j)\pi r)}{2(i+j)\pi}\bigg|_0^1\right) + \frac{1}{2i\pi}\left(-\frac{\cos(2(i-j)\pi r)}{2(i-j)\pi}\bigg|_0^1\right) = 0.\end{aligned}$$

For $i = j$, we have

$$\int_0^1 \Psi_i^c(r)\phi_j^c(r)\,dr = \frac{1}{2i\pi}\left(\int_0^1 \sin(2(i+j)\pi r)\,dr\right) = \frac{1}{2i\pi}\left(-\frac{\cos(2(i+j)\pi r)}{2(i+j)\pi}\bigg|_0^1\right) = 0.$$

Therefore $\int_0^1 \Psi_i^c(r)\phi_j^c(r)\,dr = 0$ for any given $i, j = 1, ..., K$. That is, the cosine functions also satisfy Assumption 3. ■

Proof of Theorem 7. We have

$$T(\hat{\beta}_{TOLS} - \beta_0) \Longrightarrow \Omega_{xx}^{-1/2}\left(\sum_{i=1}^K \eta_i\eta_i'\right)^{-1}\left(\sum_{i=1}^K \eta_i\psi_i\right),$$

where

$$\eta_i := \int_0^1 \phi_i(r)W_x(r)dr \text{ and } \psi_i := \sigma_{0\cdot x}\nu_i + \sigma_{x0}\Omega_{xx}^{-1/2}\xi_i$$

for

$$\xi_i := \int_0^1 \phi_i(r)dW_x(r) \text{ and } \nu_i = \int_0^1 \phi_i(r)dw_0(r).$$

Since $\psi := (\psi_1, ..., \psi_K)' \sim N(0, \sigma_0^2 I_K)$ and $\eta \perp \psi$, we can represent the limiting distribution as zero mean mixed normal distribution

$$\Omega_{xx}^{-1/2}\left(\sum_{i=1}^K \eta_i\eta_i'\right)^{-1}\left(\sum_{i=1}^K \eta_i\psi_i\right) = \Omega_{xx}^{-1/2}\left(\eta'\eta\right)^{-1}\eta'\psi \overset{d}{=} MN\left(0, \sigma_0^2\Omega_{xx}^{-1/2}(\eta'\eta)^{-1}\Omega_{xx}^{-1/2}\right)$$

as desired. ■

Proof of Proposition 9. Note that $Ec'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c = c'E\left[E(V_{TAOLS}^{-1} - V_{TOLS}^{-1})|\eta\right]c$ for any conforming vector $c \in \mathbb{R}^d$. From equation (65) and the independence between η and ξ, we have

$$c'\left[E(V_{TAOLS}^{-1} - V_{TOLS}^{-1})|\eta\right]c = \frac{c'\Omega_{xx}^{1/2}\eta'}{\sigma_{0\cdot x}}\left(\varrho^2 I_K - E[P_\xi]\right)\frac{\eta\Omega_{xx}^{1/2}c}{\sigma_{0\cdot x}}. \tag{92}$$

where the (i,j)th element of $P_\xi = \xi(\xi'\xi)^{-1}\xi'$ is

$$p_{i,j} = \xi_i' \left(\sum_{s=1}^{K} \xi_s \xi_s'\right)^{-1} \xi_j.$$

We want to show that $E[P_\xi] = d/K \cdot I_K$. That is, $E[p_{ii}] = d/K$ for $i = 1, ..., K$ and $E[p_{ij}] = E[p_{ji}] = 0$ for all $i \neq j$. Since $\xi_i \overset{i.i.d}{\sim} N(0, I_d)$, it is easy to show that

$$\xi_i' \left(\sum_{s=1}^{K} \xi_s \xi_s'\right)^{-1} \xi_i \overset{d}{=} \xi_j' \left(\sum_{s=1}^{K} \xi_s \xi_s'\right)^{-1} \xi_j \text{ for all } i, j = 1, ..., K. \tag{93}$$

So, we have $E[p_{ii}] = E[p_{jj}]$, i.e., all the diagonal elements of $E[P_\xi]$ are same. In other words, $E[p_{11}] = E[p_{22}] = ... = E[p_{KK}] = \lambda$ for some λ. This gives us

$$\lambda K = tr[EP_\xi] = E[tr\left[\xi(\xi'\xi)^{-1}\xi'\right] = d$$

and so $\lambda = E[p_{ii}] = d/K$ for $i = 1, ..., K$. For the off-diagonal elements, we note that the distribution of $p_{i,j}$ is symmetric around zero, which implies that $E[p_{ij}] = E[p_{ji}] = 0$ for all $i \neq j$. Therefore,

$$Ec'(V_{TAOLS}^{-1} - V_{TOLS}^{-1})c = \left(\varrho^2 - \frac{d}{K}\right) E \left\| \frac{\eta \Omega_{xx}^{1/2} c}{\sigma_{0\cdot x}} \right\|^2, \tag{94}$$

and this immediately leads to the desired result. ■

References

[1] Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.

[2] Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics* 77(2), 379–404.

[3] Bunzel, H. (2006). Fixed-b asymptotics in single-equation cointegration models with endogenous regressors. *Econometric Theory* 22, 743–755.

[4] Jin, S., Phillips, P.C.B., Sun, Y. (2006). A new approach to robust inference in cointegration. *Economics Letters* 91, 300–306.

[5] Hamming, R.W. (1973). Numerical Methods for Scientists and Engineers. Dover Publications, New York.

[6] Hwang, J., Sun, Y. (2016). Simple, Robust, and Accurate F and t Tests in Cointegrated Systems. Working paper, Department of Economics, UC San Diego, http://escholarship.org/uc/item/82k1x4rd

[7] Hwang, J., Sun, Y. (2017). Asymptotic F and t Tests in an Efficient GMM Setting. *Journal of Econometrics*, 198(2), 277-295.

[8] Kiefer, N.M., Vogelsang, T.J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.

[9] Lazarus, E., Lewis, D., Stock, J.H., Watson, M.W. (2016). HAR inference: kernel choice, size distortions, and power losses. Working paper, Department of Economics, Harvard University.

[10] Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59 (6), 1551–1580.

[11] Müller, U.K. (2007): "A theory of robust long-run variance estimation." *Journal of Econometrics* 141, 1331–1352.

[12] Müller, U. K. (2014). HAC corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics* 32(3), 311–322.

[13] Müller, U. K., Watson, M.W. (2013). Low-frequency robust cointegration testing. *Journal of Econometrics* 174(2), 66–81.

[14] Müller, U.K., Watson, M.W. (2016). Low frequency econometrics. Working paper, Department of Economics, Princeton University.

[15] Phillips, P.C.B. (1991a). Spectral regression for cointegrated time series. *Nonparametric and Semiparametric Methods in Economics and Statistics.* Edited by Barnett, W., Powell, J., Tauchen, G., Cambridge University Press, New York, pp. 413–435.

[16] Phillips, P.C.B. (1991b). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.

[17] Phillips, P.C.B. (2005a). Automated discovery in econometrics. *Econometric Theory* 21(1), 3–20.

[18] Phillips, P.C.B. (2005b). HAC estimation by automated regression. *Econometric Theory* 21(1), 116–142.

[19] Phillips, P.C.B. (2014). Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics* 178(2), 210–224.

[20] Phillips, P.C.B., Durlauf, S.N. (1986). Multiple regression with integrated processes. *Review of Economic Studies* 53, 473–496.

[21] Phillips, P.C.B., Hansen, B.E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.

[22] Phillips, P.C.B., Liao, Z. (2014). Series estimation of stochastic processes: recent developments and econometric applications. *Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, edited by J.S. Racine, L. Su, and A. Ullah.

[23] Phillips, P.C.B., Loretan, M. (1991). Estimating long run economic equilibria. *Review of Economic Studies* 58, 407–436.

[24] Phillips, P.C.B., Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica* 58 (1), 165–193.

[25] Saikkonen, P. (1991). Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7, 1–21.

[26] Shintani, M. (2001). A simple cointegrating rank test without vector autoregression. *Journal of Econometrics* 105, 337–362.

[27] Stock, J.H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica* 55, 1035–1056.

[28] Stock, J.H., Watson, M.W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.

[29] Stoica, P., Moses, R. (2005). *Spectral Analysis of Signals.* Pearson: New Jersey.

[30] Sun, Y. (2006). "Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise." Working paper, Department of Economics, UC San Diego.

[31] Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics* 164(2), 345–66.

[32] Sun, Y. (2013). A heteroskedasticity and autocorrelation robust F test using orthonormal series variance estimator. *Econometrics Journal* 16(1), 1–26.

[33] Sun, Y. (2014a). Let's fix It: fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust inference. *Journal of Econometrics* 178(3), 659–677.

[34] Sun, Y. (2014b). Fixed-smoothing asymptotics in a two-step GMM framework. *Econometrica* 82(6), 2327–2370.

[35] Sun, Y. (2014c). Fixed-smoothing asymptotics and asymptotic F and t tests in the presence of strong autocorrelation. *Advances in Econometrics* 33, Essays in Honor of Peter C. B. Phillips, 23–63.

[36] Sun, Y., Phillips, P.C.B., Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* 76(1), 175–94.

[37] Sun, Y., Kim, M.S. (2012). Simple and powerful GMM over-identification tests with accurate size. *Journal of Econometrics* 166(2), 267–281.

[38] Sun, Y., Kim, M.S. (2015). Asymptotic F test in a GMM framework with cross sectional dependence. *Review of Economics and Statistics* 97(1), 210–223.

[39] Thomson, D.J. (1982). Spectrum estimation and harmonic analysis. *IEEE Proceedings* 70, 1055–1096.

[40] Vogelsang, T.J., Wagner, M. (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics* 178(2), 741–760.

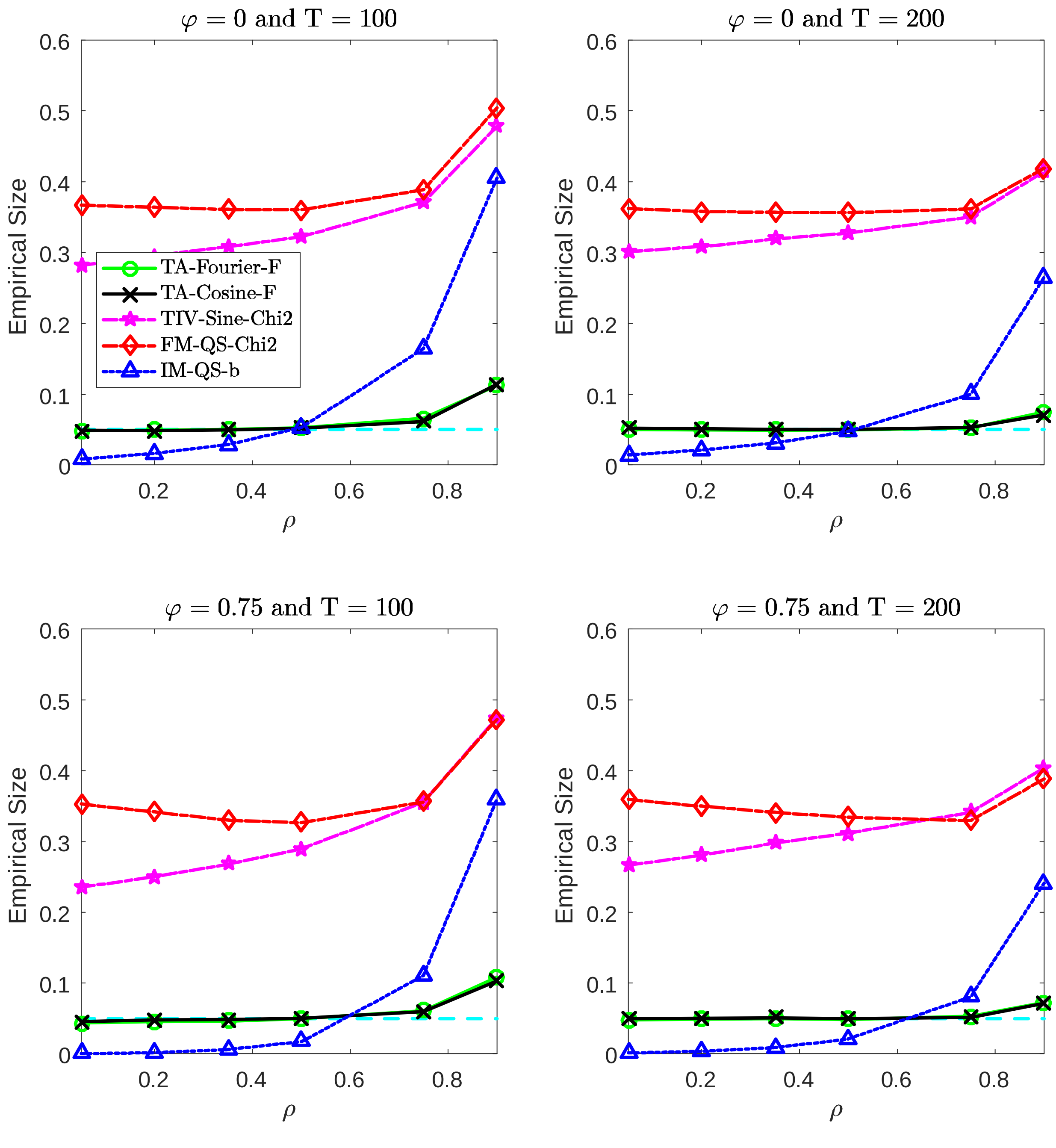

Figure 1: Empirical size of 5% fixed-smoothing tests (TA-Fourier-F, TA-Cosine-F, IM-QS-b) and chi-squared tests (TIV-Sine-Chi2, FM-QS-Chi2) with $K = 8$ and comparable b ($b = 0.13$)

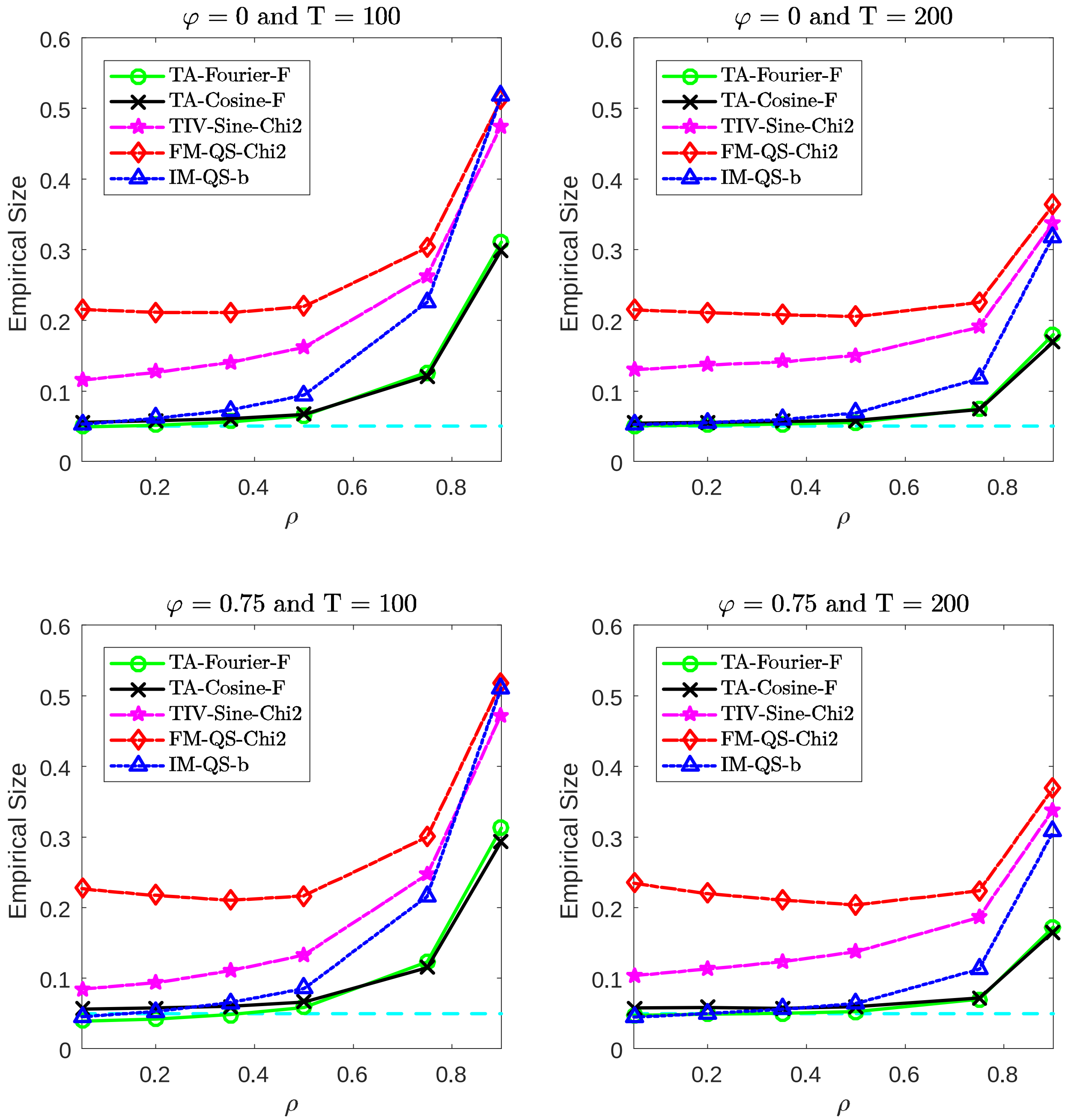

Figure 2: Empirical size of 5% fixed-smoothing tests (TA-Fourier-F, TA-Cosine-F, IM-QS-b) and chi-squared tests (TIV-Sine-Chi2, FM-QS-Chi2) with $K = 16$ and comparable b ($b = 0.06$)

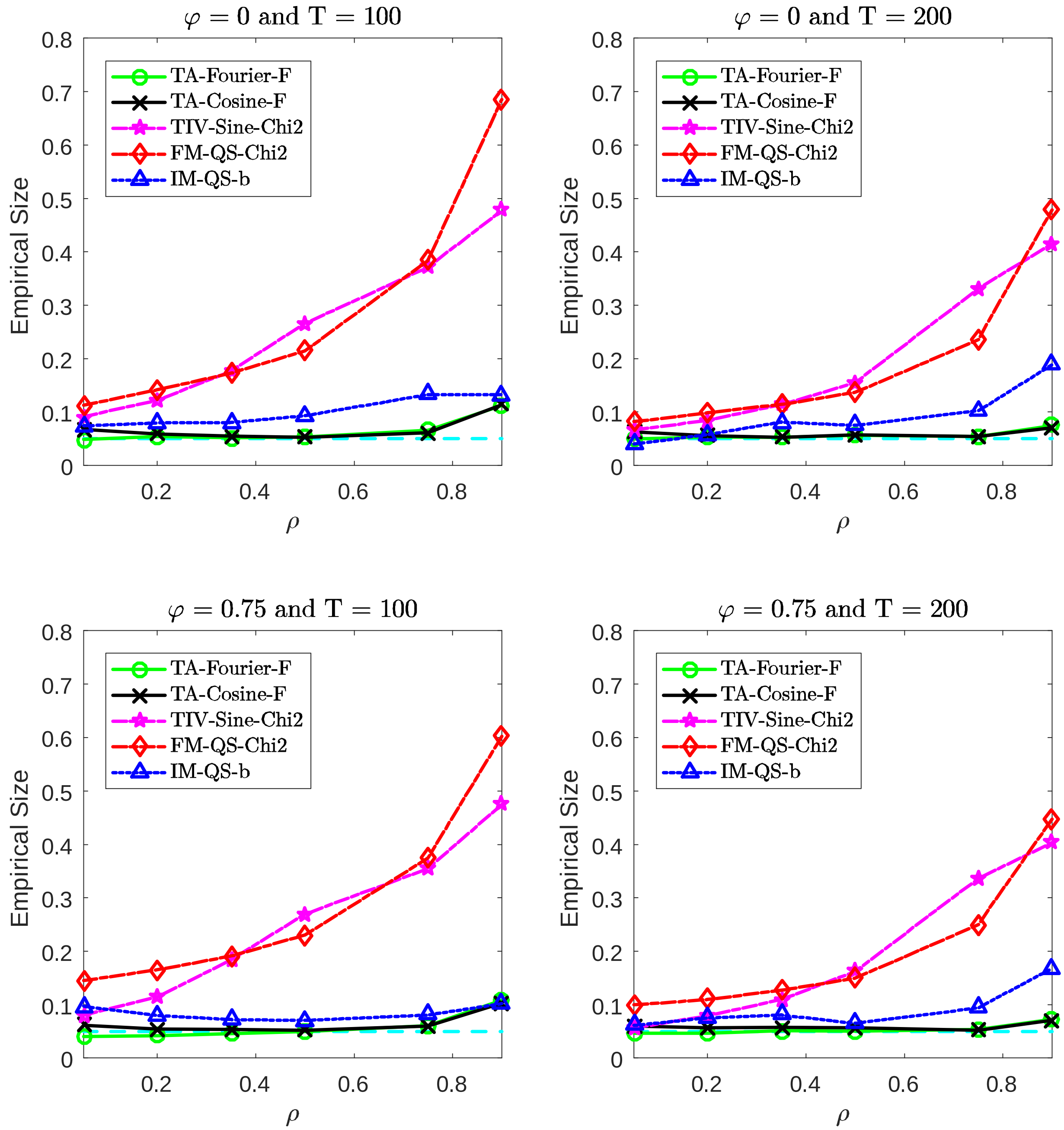

Figure 3: Empirical size of 5% fixed-smoothing tests (TA-Fourier-F, TA-Cosine-F, IM-QS-b) and chi-squared tests (TIV-Sine-Chi2, FM-QS-Chi2) with data-driven smoothing parameters

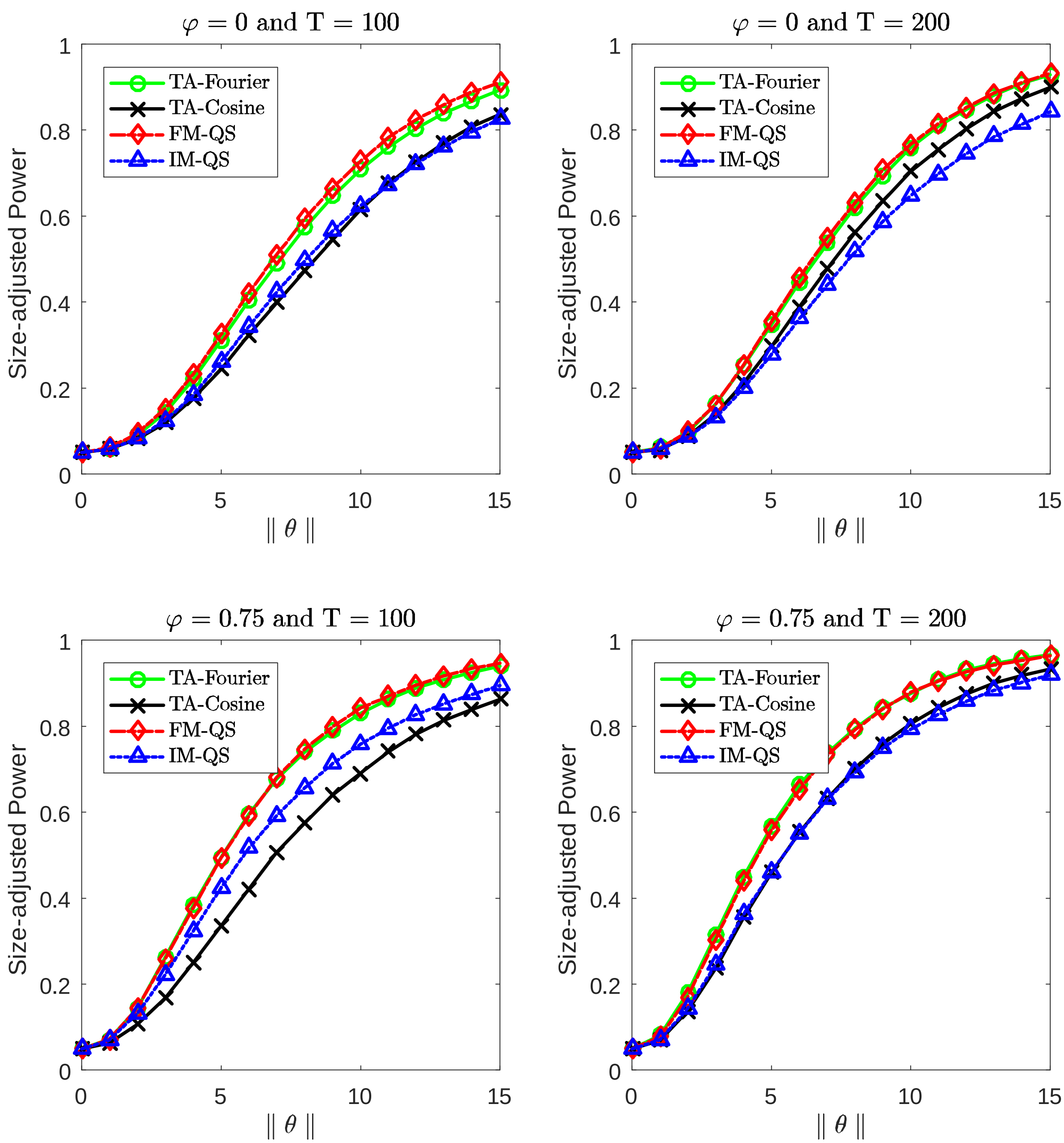

Figure 4: Size-adjusted power of different tests with data-driven smoothing parameters when $\rho = 0.05$

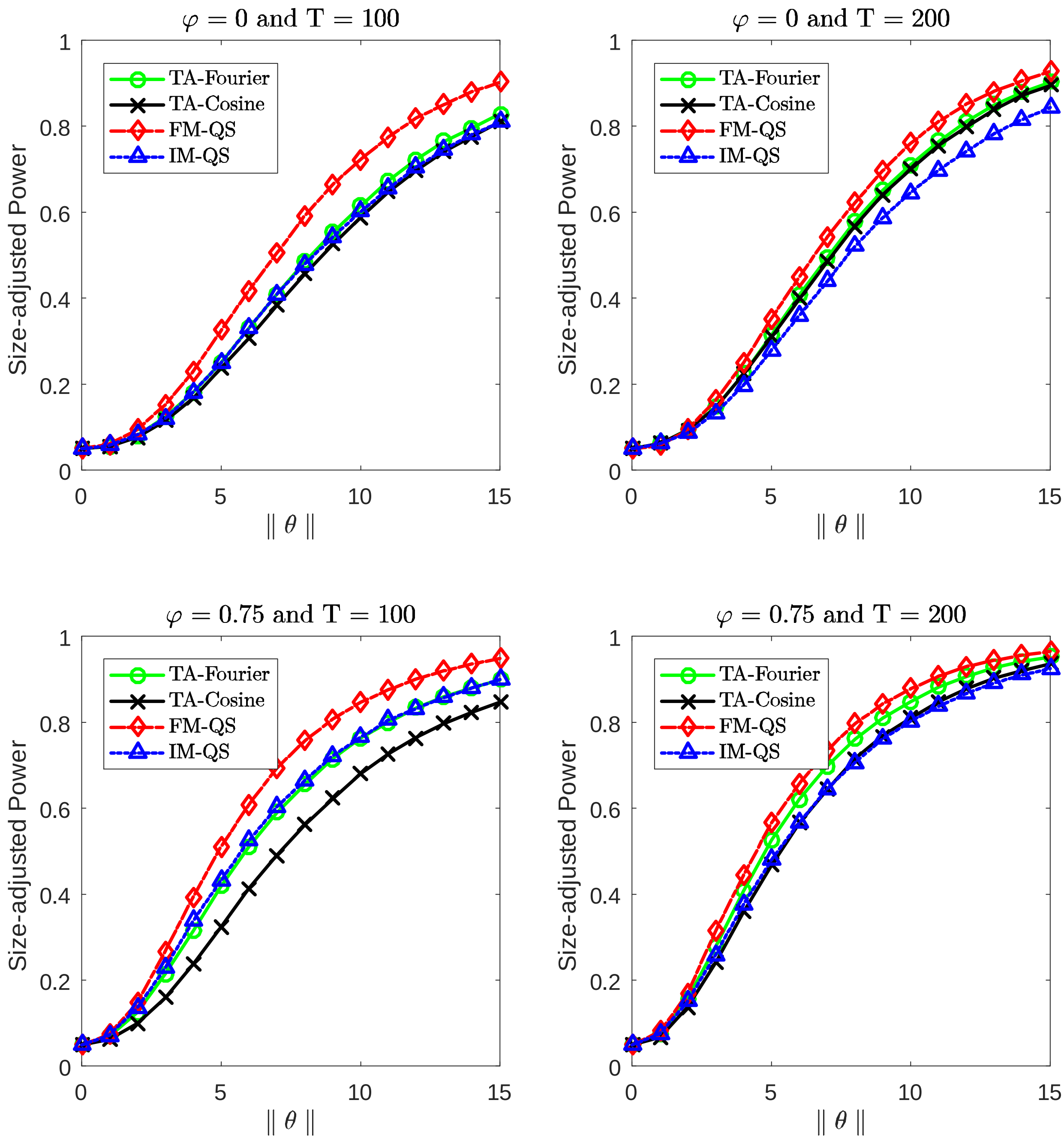

Figure 5: Size-adjusted power of different tests with data-driven smoothing parameters when $\rho = 0.35$

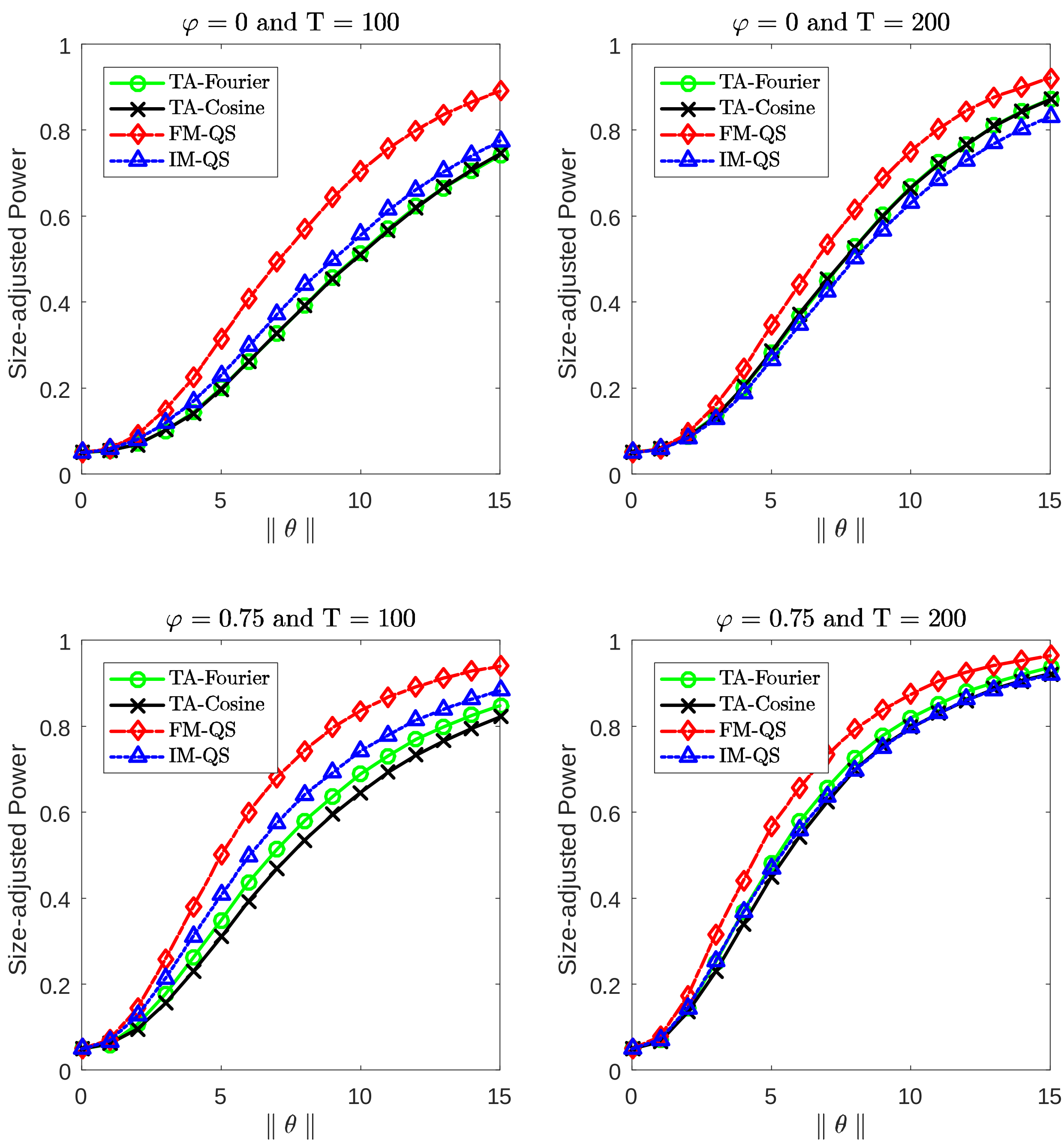

Figure 6: Size-adjusted power of different tests with data-driven smoothing parameters when $\rho = 0.50$

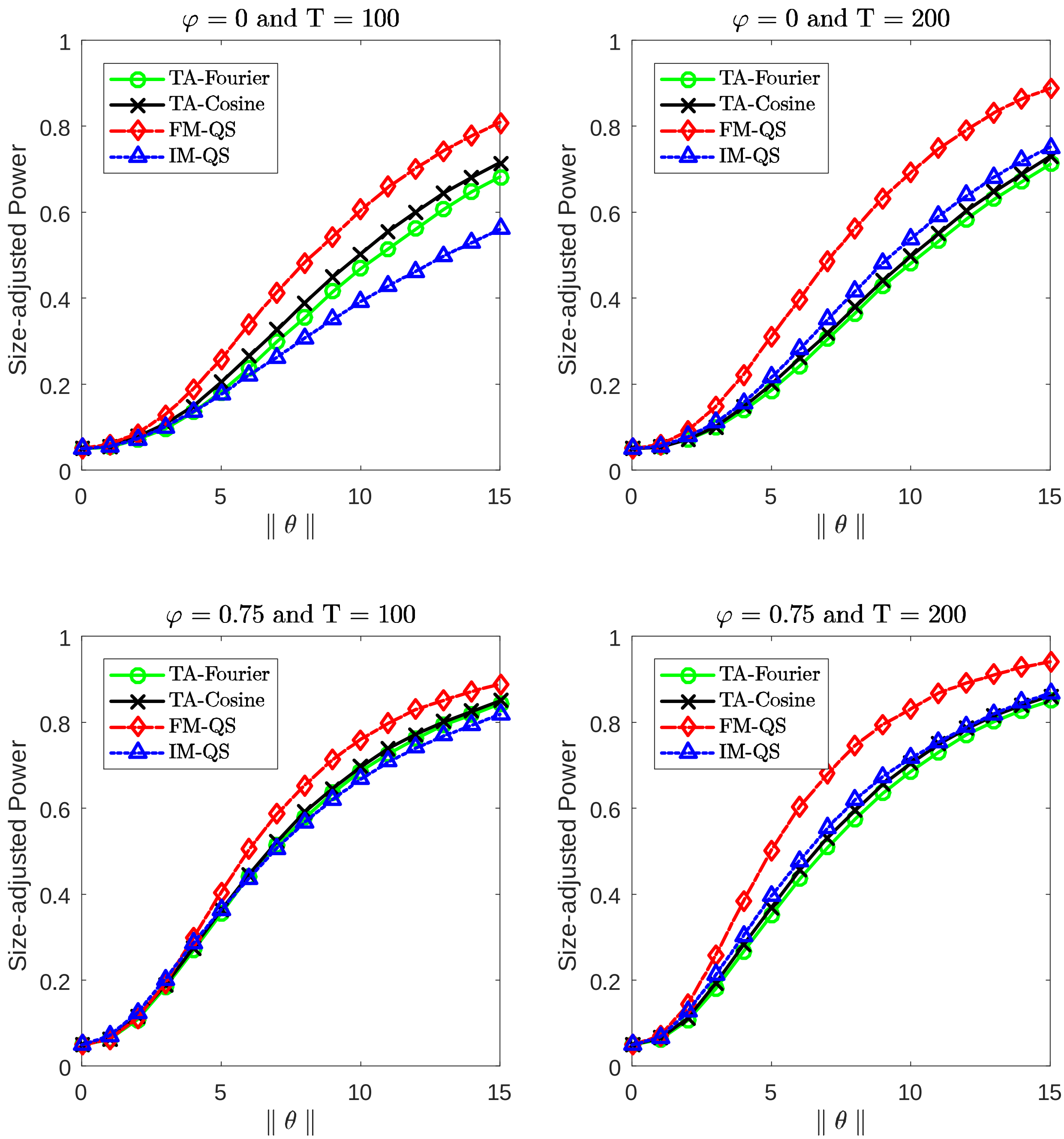

Figure 7: Size-adjusted power of different tests with data-driven smoothing parameters when $\rho = 0.75$

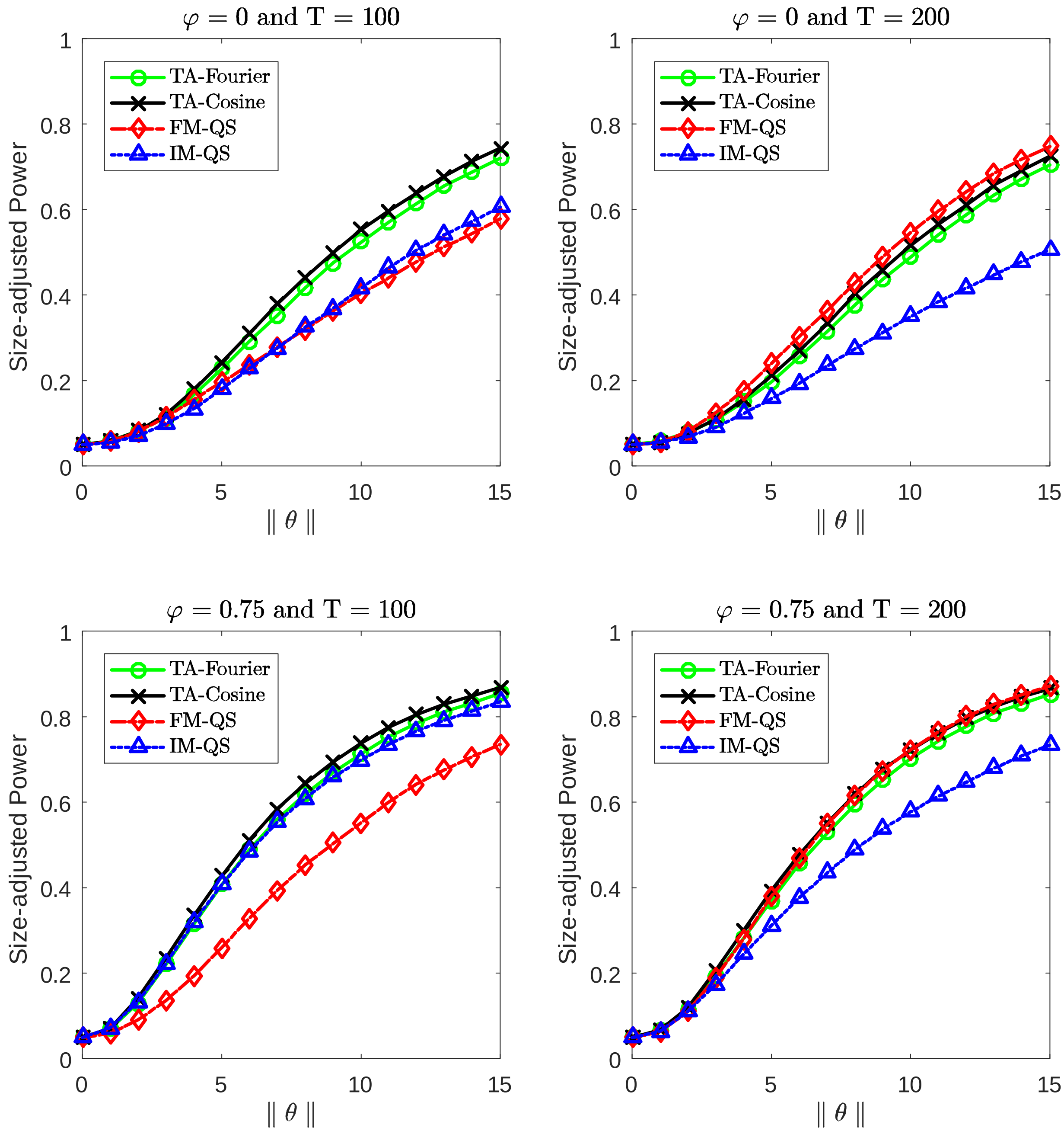

Figure 8: Size-adjusted power of different tests with data-driven smoothing parameters when $\rho = 0.90$